

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

**Scienze Biotecnologiche e Farmaceutiche**

Ciclo XXX

**Settore Concorsuale: 03/D1**

**Settore Scientifico Disciplinare: CHIM/08**

**Fully Flexible Binding of Taxane-Site Ligands to Tubulin  
via Enhanced Sampling MD Simulations**

**Presentata da:** Dario Gioia

**Coordinatore Dottorato**  
Prof. Santi Mario Spampinato

**Supervisore**  
Prof. Maurizio Recanatini

**Co-Supervisore**  
Dott. Federico Falchi

**Esame finale anno 2018**

*“Non troverai mai la verità,  
se non sei disposto ad accettare anche  
ciò che non ti aspettavi di trovare.”*

# Table of Contents

<b>ABSTRACT.....</b>	<b>1</b>
<b>LIST OF ACRONYMS .....</b>	<b>2</b>
<b>LIST OF FIGURES .....</b>	<b>5</b>
<b>1. INTRODUCTION .....</b>	<b>10</b>
1.1. The Traditional Approach to Molecular Docking .....	12
1.1.1. Posing.....	13
1.1.2. Scoring .....	14
1.1.3. Combining Docking and Molecular Dynamics Simulations .....	16
1.2. Towards Dynamic Docking.....	19
1.2.1. Unbiased-MD Approaches.....	20
1.2.1.1. Long Plain MD Simulations .....	20
1.2.1.2. Adaptive Sampling Methods .....	22
1.2.2. Enhanced Sampling Methods in Molecular Dynamics .....	25
1.2.2.1. Methods Relying on CV(s).....	26
1.2.2.2. Methods Relying on Tempering .....	28
1.3. The MD-Binding Approach .....	30
1.3.1. On the Choice of the “Right” CV(s) .....	30
1.3.2. An Adaptive Bias for Ligand Binding .....	31
<b>2. TUBULIN AS A TEST CASE .....</b>	<b>34</b>
2.1. Microtubules Structure and Functions .....	34
2.2. MTs Dynamic Behavior.....	36
2.3. MTs bricks: $\alpha$ and $\beta$ -tubulin .....	39
2.4. Tubulin as Target in Cancer Therapy .....	42
2.5. Chemical Compounds that Interact with Tubulin .....	43

2.5.1.	The Taxane-Site .....	46
<b>2.6.</b>	<b>Aims of the Thesis .....</b>	<b>49</b>
<b>3.</b>	<b>MATERIAL AND METHODS .....</b>	<b>51</b>
<b>3.1.</b>	<b>MD Simulations Setup .....</b>	<b>51</b>
3.1.1.	System Preparation .....	51
3.1.2.	Plain MD Simulations .....	52
<b>3.2.</b>	<b>Binding Pocket Analysis .....</b>	<b>52</b>
<b>3.3.</b>	<b>MD-Binding Protocol Setup .....</b>	<b>53</b>
3.3.1.	Ligands Parametrization .....	53
3.3.1.1.	Epothilone A .....	53
3.3.1.2.	Discodermolide .....	53
3.3.1.3.	Taxol .....	53
3.3.2.	Ligand(s) Positioning .....	54
3.3.3.	Systems Preparation and Equilibration .....	54
3.3.4.	MD-Binding Simulations .....	54
3.3.5.	Molecular Docking .....	56
3.3.5.1.	Traditional Static Docking .....	56
3.3.5.2.	Induced-Fit Docking .....	56
3.3.6.	Scoring of MD-Binding Poses .....	56
3.3.6.1.	Rescoring with MM/GBSA .....	56
3.3.7.	Figure Preparation .....	57
<b>4.</b>	<b>RESULTS AND DISCUSSION .....</b>	<b>58</b>
<b>4.1.</b>	<b>Pocket Behavior Investigations .....</b>	<b>58</b>
4.1.1.	Model Selection .....	58
4.1.2.	Classical MD-Simulation .....	61
4.1.3.	Tracking the Binding Pocket .....	64
<b>4.2.</b>	<b>Considerations on the MD-Binding Protocol Setup .....</b>	<b>71</b>
<b>4.3.</b>	<b>Protocol Validation .....</b>	<b>74</b>



4.3.1.	Traditional Docking Procedures.....	74
4.3.2.	The Two Epothilone A Binding Modes at Test.....	75
4.3.3.	Dynamic Docking of Epothilone A.....	77
4.3.3.1.	Posing.....	77
4.3.3.2.	The importance of the His227 .....	83
4.3.3.3.	Scoring .....	84
<b>4.4.</b>	<b>Dynamic Docking of Discodermolide and Taxol.....</b>	<b>87</b>
4.4.1.	Discodermolide .....	87
4.4.2.	Taxol .....	90
<b>5.</b>	<b>CONCLUSIONS AND PERSPECTIVES .....</b>	<b>93</b>
	<b>APPENDIX: RESIDUE NUMBERING CORRESPONDENCE .....</b>	<b>95</b>
	<b>REFERENCES.....</b>	<b>97</b>

# ABSTRACT

Microtubules (MTs) are cytoskeleton components involved in a plenty of cellular functions such as transport, motility, and mitosis. Being polymers made up of  $\alpha/\beta$ -tubulin heterodimers, in order to accomplish these functions, they go through large variations in their spatial arrangement switching between polymerization and depolymerization phases.

Because of their role in cellular division, interfering with MTs dynamic behavior has been proven to be suitable for anticancer therapy as tubulin-binding agents induce mitotic arrest and cell death by apoptosis. However, how microtubule-stabilizing agents like taxane-site ligands act to promote microtubule assembly and stabilization is still argument of debate.

As in the case of tubulin, traditional docking techniques lack the necessary capabilities of treating protein flexibility that are central to certain binding processes. For this reason, the aim of this project is to put in place a protocol for dynamic docking of taxane-site ligands to  $\beta$ -tubulin by means of enhanced sampling MD simulation techniques.

Firstly, the behavior of the binding pocket has been investigated with classical MD simulations. It has been observed that the most flexible part of the taxane site is the so-called “M-loop”, the one involved into the lateral associations of tubulin heterodimers and that is supposed to be stabilized by taxane-site ligands.

Secondly, the protocol for the dynamic docking has been put in place using the MD-Binding technique developed by BiKi Technologies. It showed to be effective in reproducing the binding mode of epothilone A and discodermolide as in their X-ray crystal structures.

Finally, the protocol has been tested against paclitaxel, a drug for which no X-ray crystal structure is currently available.

These results showed the potential of such an approach and strengthen the belief that in the future dynamic docking will replace traditional static docking in the drug discovery and development process.

## List of Acronyms

ACE	Angiotensin-Converting Enzyme
ACh	Acetylcholine
aMD	Accelerated Molecular Dynamics
BD	Brownian Dynamics
BS-MetaD	Bias-Exchange-metadynamics
CADD	Computer-Aided Drug Design
CDK2	Cyclin-Dependent Kinase 2
Cryo-EM	Cryo-Electron Microscopy
CV	Collective Variable
DADME	DADMe-immucillin-H [4'-deaza-1'-aza-2'-deoxy-1'-(9-methylene)-Immucillin-H]
DDM	(+)-Discodermolide
EP	Epothilone A
E-site	Exchangeable Nucleotide Site
FEP	Free Energy Perturbation
GDP	Guanosine Diphosphate
GMPCPP	Guanosine-5'-[( $\alpha,\beta$ )-methyleno]triphosphate
GPCR	G-Protein Coupled Receptor
GPU	Graphics Processing Unit
GTP	Guanosine Triphosphate
HDX-MS	Hydrogen-Deuterium Exchange coupled to Mass-Spectroscopy
H-REMD	Hamiltonian Replica-Exchange Molecular Dynamics
HTD	High Throughput Docking

HTMD	High-Throughput Molecular Dynamics
HTS	High-Throughput Screening
LIE	Linear Interaction Energy
MAPs	Microtubule-Associated Proteins
McMD	Multicanonical Molecular Dynamics
MD	Molecular Dynamics
MetaD	Metadynamics
MHC	Major Histocompatibility Complex
MM/GBSA	Molecular Mechanics-Generalized Born Surface Area
MM/PBSA	Molecular Mechanics-Poisson-Boltzmann Surface Area
MSM	Markov State Model
MT	Microtubule
MTOC	Microtubule Organizing Center
N-site	Nonexchangeable Nucleotide Site
PBC	Periodic Boundary Conditions
PNP	Purine Nucleoside Phosphorylase
R	RB3-SLD – Stathmin-Like Domain of the neural protein RB3
REMD	Replica-Exchange Molecular Dynamics
RMD	Reconnaissance Metadynamics
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SBDD	Structure-Based Drug Design
SES	Solvent Excluded Surface
SMILES	Simplified Molecular-Input Line-Entry System
STD	Saturation Transfer Difference (Analysis)

SuMD	Supervised Molecular Dynamics
T	Tubulin
TAX	Taxol
TI	Thermodynamic Integration
TTL	Tubulin-Tyrosine Ligase
US	Umbrella Sampling
VS	Virtual Screening
$\gamma$ TuRC	$\gamma$ -Tubulin Ring Complex

## List of Figures

- Figure 1 – MD simulations are employed for generating a conformational ensemble of protein structures. A number of snapshots are selected by cluster analysis and then docking is performed against these multiple rigid receptor conformations to evaluate protein flexibility..... 17
- Figure 2 – MD simulations run after docking with the purpose of refining and rescoreing the obtained binding modes. .... 18
- Figure 3 - In MetaD, a bias is applied to a defined CV ( $q$ ) in order to fill the underlying free energy ( $F(q)$ ) and discouraging the system to sample already sampled states. .... 27
- Figure 4 – Among the methods that do not rely on CVs, REMD simulates independently several replicas of the system at different temperature ( $T$ ) giving them the possibility to exchange coordinates at regular intervals. Background color indicates the Temperature. Rep1-4 are indicated with different colors. .... 29
- Figure 5 – Structure of a 15 protofilaments MT generated from the cryo-electron microscopy (cryo-EM) in complex with the *Drosophila melanogaster* kinesin-13 KLP10A (PDB ID: 3J2U - 10.8 Å) after the removal of the kinesin protein. From the top view (on the left) it is possible to observe the internal lumen that is characteristic of MTs. The way tubulin dimers are assembled generates a polarity in MT structure (side view, on the right).  $\beta$ -tubulin monomers are depicted in blue surfaces,  $\alpha$ -tubulin in cyan..... 35
- Figure 6 – Tubulin heterodimers.  $\beta$ -tubulin monomers are depicted in blue surfaces,  $\alpha$ -tubulin in cyan. .... 35
- Figure 7 – Spontaneous MTs nucleation and growth. The figure depicts the whole process, starting from a solution of purified tubulin heterodimers with GTP and  $Mg^{2+}$  (on the left) to the complete assembly (on the right). The red line indicates the seam, i.e. the discontinuity in the lattice due to the vertical offset between tubulin monomers. .... 36
- Figure 8 – The influence of nucleotide hydrolysis on tubulin dimer curvature and MTs treadmilling. GTP-bound tubulin islands (in red, at the center of the MT) may act as a trigger for switching among depolymerizing/polymerizing phases. .... 38
- Figure 9 - Microtubule dynamic instability. GTP-bound tubulin assembles at the microtubule plus end. After GTP hydrolysis occurs, the microtubule becomes unstable and

depolymerizes at the same end. When GDP is exchanged with GTP in the free tubulin dimers, the cycle can begin again. ....38

Figure 10 – Tubulin in complex with Taxol (PDB ID: 1JFF – 3.5 Å) and GTP (in the  $\alpha$ -tubulin, down in the picture) and GDP ( $\beta$ -tubulin, up). Functional domains, as reported in [118], are colored differently: N-terminal domain in yellow; drug-binding domain in slate blue; C-terminal domain in light magenta. ....40

Figure 11 – Structural differences between heterodimers in straight polymerized protofilaments as within microtubules (on the left), and in curved complexes as it happens upon binding of the depolymerizing protein stathmin (on the right). ....41

Figure 12 – The mitotic spindle [138]. ....42

Figure 13 – The five different binding sites in the  $\beta$ -tubulin subunit. Colchicine (yellow dots) binds at the level of the intradimer interface. Vinca alkaloids (dark green dots), such as vinblastine, bind at the interdimer interface, so as maytansine-site ligands (purple dots). MSAs, such as taxane-site ligands (i.e. Epothilone A, white dots) and peloruside A (lime green dots) bind close to the lateral interface between protofilaments. ....44

Figure 14 – Location of the two MSA sites within the MT shown from the plus-end perspective. Taxol binds on  $\beta$ -tubulin (dark grey spheres) facing the lumen of the MT. Peloruside also binds at the interface between protofilaments, but facing the external part of the MT. ....46

Figure 15 – The tubulin taxane-site when Taxol is bound as in the EC crystal structure (PDB ID: 1JFF). ....47

Figure 16 – Comparison between the originally proposed EP binding mode (on the left, PDB ID: 1TVK) and the recently one obtained through X-ray (on the right, PDB ID: 4I50). It is possible to note how the EC model lacks the necessary resolution to unambiguously assign the position of EP atoms, as is the case of the X-ray structure. ....48

Figure 17 – Zampanolide binding mode as in its X-ray crystal structure (PDB ID: 4I4T – 1.8 Å). The ligand is covalently bound to the His 229 of Helix H7. Other than two hydrogens bond with the backbone of Thr 276, it is possible to identify many hydrophobic interactions. ....59

Figure 18 – The selected protein structure for this work. Chains removed from the PDB 4I4T are transparent, as it is the ligand zampanolide in the $\beta$ -tubulin taxane-site. $\beta$ -tubulin is colored in marine blue, $\alpha$ -tubulin in light cyan.....	60
Figure 19 - Calculated RMSD of the $\alpha\beta$ -tubulin heterodimer (A), $\alpha$ -tubulin (B), and $\beta$ -tubulin (C) subunits from their initial X-ray structure plotted as a function of time. The lightest lines represent the effective sampling of the RMSD during the entire simulation, whereas the darkest ones are smoothed in order to cut-off the noise. ....	62
Figure 20 - Calculated RMSD of the C-terminal domains (CTDs) of both subunits from their initial X-ray structure plotted as a function of time. The red lines represent the RMSD computed taken into account the contribution of the entire CTD, while the blue ones have been obtained removing the last disordered residues. ....	62
Figure 21 – Calculated RMSD of the C $\alpha$ carbons of the three domains from their initial X-ray structure plotted as a function of time. On the left, the RMSD associated to the $\alpha$ -tubulin subunit. On the right, the one of the $\beta$ -tubulin monomer. ....	63
Figure 22 – Calculated RMSF of the C $\alpha$ carbons of the $\beta$ -tubulin subunit during the simulation. For clarity reasons, the data coming from the most flexible C-terminal tails (residues 426-431) has been removed. ....	64
Figure 23 – Marge matrix of the 36 identified pockets. Values from 0 to 40 represent the merging probability between two pockets. For clarity reason, values over 40 have been removed. ....	67
Figure 24 - 3D network graph of pocket communication. The nodes represent pockets, and the edges indicate communication between two pockets. The thickness of each edge is directly proportional to the frequency of the merging events. Only pockets with a persistency greater than 40 % are shown.....	68
Figure 25 - Time persistency of residues that define the main taxane pocket (blue bars, pocket ID 4) and its correlated pocket (green bars, pocket ID 2). Only residues with a persistency greater than 30% are shown.....	69
Figure 26 – Volume over time (expressed as frame number) of the two taxane-site pockets (ID 2 and 4). Light colored lines represent the effective sampling of the volume during the simulation, whereas the dark colored ones represent their approximation with a Bezier curve. ....	70



Figure 27 – MD-Binding workflow.....	71
Figure 28 - Pocket volume (showed in blue surface) at different times during the simulation generated by NanoShaper. On the left, the 875 <sup>th</sup> frame used for generating the pocket entrance file, and, on the right, the last frame of the simulation, where the taxane-site pocket is split in two different pockets, therefore pocket ID 4 volume is much smaller. ...	73
Figure 29 - Calculated RMSD of the ligand heavy atoms from their initial crystallographic structure plotted as a function of time, after fitting on binding site C $\alpha$ carbons.....	75
Figure 30 - Calculated RMSF of the C $\alpha$ carbons of the $\beta$ -tubulin subunit during the simulations of the two complexes. For clarity reasons, the data coming from the most flexible C-terminal tails (residues 426-431) has been removed. ....	76
Figure 31 – Reference setup 1 of Table 4. Attractive residues are shown in blue, switching-off residues in magenta. ....	77
Figure 32 - Reference setups 5-12 of Table 4. Attractive residues (Leu655, His667, Phe710, Pro712, Thr714, and Gln719) are shown in blue, the switching-off residue Pro712 in magenta.....	79
Figure 33 – Comparison between the crystallographic binding mode of epothilone A (top) and the one obtained by means of MD-Binding (bottom). Residues with a different conformation (Thr714 and Gln719) are highlighted in orange.....	80
Figure 34 – The correctly reproduced binding mode of epothilone A with MD-Binding. Reference setup 16 of Table 4, Run ID 5. For comparison, see Figure 33 (top). ....	82
Figure 35 – Evolution of the bias (grey area) for two different MD-Binding setups. The green line represents the distance between the ligand and the switching-off residue Pro712 monitored along the MD simulations. The red line is the imposed threshold value of 4 Å that controls the bias switch-off.....	83
Figure 36 – The orientation of the His227 is important for the correct positioning of epothilone A macrolidic core. His227 is colored in orange when its position, as observed in many MD-Binding runs, does not allow the proper positioning of epothilone A. In green, as reported in the crystallographic structure (PDB ID: 4I50). ....	84

Figure 37 – Comparison between the binding mode of DDM (on the left) and EP (on the right) as reported in their respective X-ray structures. The highly flexible M-loop, not interacting with DDM, is not resolved.....	88
Figure 38 - Calculated RMSD of DDM and pocket backbone to the X-ray structure (PDB ID: 5LXT) plotted as a function of time. ....	88
Figure 39 – The correctly reproduced binding mode of DDM with our MD-Binding protocol ( Reference setup 14 of Table 4, Run ID 4). For comparison, see Figure 37 (left). The induced $\alpha$ -helix is still well folded. ....	89
Figure 40 – Comparison of bias and distance evolution during the course of TAX MD-Binding simulations. On the left, a simulation in which TAX did not make contact with the switching-off residue Pro272; on the right, the bias is completed switched off at about 7 ns, and the simulation reached a plain MD-like status.....	91
Figure 41 – The taxol obtained discordant binding modes. The selected switching-off residue Pro272 is shown in orange. ....	91
Figure 42 – The taxol baccatin core highlighted in red. ....	92

# 1. Introduction

Proteins are inherently flexible objects, capable of undergoing large motions in their 3D structure and adopting different conformations in order to exert their functions. Without this capability, almost all the processes crucial for cellular life like cytoplasmic organization, metabolism, enzymatic activity, and signal transduction could not be possible. This is mostly related to the plethora of partners with whom proteins interact, ranging from ions and substrate molecules to small peptides and other proteins. The mechanism by which proteins identify their partner(s) is at the foundation of their biological activity and understanding the driving forces of this process is mandatory to design new drugs able to inhibit or enforce a specific biological process.

When enough structural data is available, computational techniques applied to the pharmaceutical and chemical fields, generally referred as computer-aided drug design (CADD) [1], showed to be effective throughout the entire drug discovery and development process. Following this process, CADD can be useful from the initial rational identification of novel compounds showing some activity against a biological target (hit compounds), in the subsequent optimization so as to increase their potency and physicochemical properties *in vitro* (lead compounds), to the final improvement in efficacy and pharmacokinetic properties [2] *in vivo* (drugs).

Conventionally, the search of hit compounds relied mostly on expensive and time-consuming biological assays (such as high-throughput screening, HTS [3]). Besides, the chemical space that can be screened is limited to the compounds that are present in a particular library owned by a company or by an academic institution. On the contrary, the availability of virtual libraries let the researchers extend their studies to compounds that do not necessarily exist physically but that can be readily obtained through purchase or synthesis [4] and, even further, to that chemical space that has not been explored yet [5], nor chemically neither computationally.

However, it has to be stressed that one cannot consider CADD as an alternative to the conventional (HTS) assays. Together, the two share strengths and weaknesses and should be used in combination so as to improve efficacy reducing the waste of resources in synthesis and screening. Taking into account the cases in which CADD helped in the discovering of chemical compounds nowadays used in therapy, some examples can be the ACE inhibitor Captopril (Bristol Myers-Squibb) and the renin inhibitor Aliskiren (Rasilez, Novartis) used for

the treatment of hypertension or the carbonic anhydrase inhibitor Dorzolamide (Merk) used against glaucoma eye pathologies [6].

In the subsequent pages, molecular docking, one of the most useful application of CADD techniques, will be introduced along with the reasons that have made it so popular and the reasons why a paradigm shift to dynamic docking is needed in order to deal with protein flexibility.

## 1.1. The Traditional Approach to Molecular Docking

When it comes to discovering novel hit compounds (i.e. compounds that show a first weak affinity with their target), virtual screening (VS) by molecular docking is still the method of election in structure-based drug discovery (SBDD) [7,8]. Molecular docking is an *in silico* technique that aims at predicting the binding mode of a specified compound within the active site of the protein of interest [9]. Thanks to the constant growth of available protein structures [10] and to the improvement in algorithms and hardware architectures, molecular docking and its fast implementation, high-throughput docking (HTD), has proven to be an invaluable tool for drug discovery campaigns [6,11]. It allows the rapid screening of large virtual libraries of compounds in a fraction of the time required by *in vitro* studies. Accordingly, only compounds that are likely to have an action against the target of interest are selected and move forward to biological tests, thus saving costs and time. Therefore, the optimization of the speed in VS approaches is crucial. In order to do so, simpler scoring functions are used and some features, necessary for a better accuracy, are omitted or underestimated.

This brings out a number of problematics:

- Protein flexibility is not treated extensively;
- Water molecules are treated only implicitly;
- Only a static picture of the binding process is provided.

As a consequence, especially when dealing with cases in which large target rearrangements are observed upon ligand binding [12] or when water molecules are crucial for ligand-target complexes, molecular docking shows its weaknesses. Therefore, a reliable estimation of key thermodynamic observables and kinetic quantities is not allowed. From this point of view, it is worth noting that binding kinetics has emerged as a more reliable predictor of *in vivo* drug efficacy than affinity [13,14] but the static picture provided by molecular docking cannot be used to estimate kinetic quantities.

Back in time, the first attempts of molecular docking were reported by Kuntz and co-workers in 1982 [9]. Initially, both ligand and protein were treated as rigid bodies but, nowadays, almost all docking programs have implemented proprietary algorithms to treat the ligand as a full flexible entity. For some of them, a partial flexibility for the receptor can be incorporated too [15,16]. It is known, in fact, that residues involved in the catalytic

mechanism of enzymes often show some grade of flexibility [17] and different side chains conformations must be considered.

Briefly, a typical docking algorithm can be divided into two distinct steps, the posing phase and the scoring phase [18].

### **1.1.1. Posing**

The first one, the posing phase, is entrusted to a searching algorithm, which can be either stochastic or deterministic. It takes care of both roto-translational and internal (usually conformational) degrees of freedom of the ligand in order to generate a set of reasonable conformations at the receptor's binding site. For practical reasons, mostly due to limited computational resources, the early docking implementations commonly dealt with a single rigid conformation of the target. However, as already stated, proteins are intrinsically dynamic, and it is not wise to neglect their flexibility.

Among the recent interpretations of protein-ligand recognition mechanism, it is worth mentioning the induced-fit model proposed by Koshland in 1958 [19]. According to it, there is a mutual adaptation of the two partners upon binding in order to maximize favorable contacts and strengthen the interaction. Likewise, another interpretation of binding, the conformational selection or population shift model [20,21], thinks of proteins as a conformational ensemble. Only upon binding, one of the available conformations is thermodynamically selected, resulting in a population shift to a new equilibrium with respect to the one in the apo form [22].

Nevertheless, it is clear that receptor's degrees of freedom should be introduced in standard docking procedures and a number of strategies have been implemented trying to mimic these recognition mechanisms. Unfortunately, due to the computational complexity of the problem, receptor and ligand degrees of freedom cannot be sampled simultaneously and the above strategies rather than a solution to the problem appear to be as a workaround. These strategies can be classified as either single-structure methods or ensemble methods.

Single-structure methods resemble the induced-fit model in which the binding pocket (or, at least, a number of user-preselected residue side-chains) is perturbed on-the-fly after ligand binding conformational search. As expected, no large perturbation in protein structure is allowed [23]. Alternatively, in ensemble methods, a series of independent docking procedures are run on an ensemble of previously generated protein conformations (either

experimentally or theoretically by means of other sampling techniques), in a similar fashion to the population shift model. Anyway, the results obtained by induced-fit approaches and the information provided by multiple protein conformations must be properly handled to be effectively exploited [24–26].

### 1.1.2. Scoring

The second phase of a docking protocol is mainly related to the evaluation of the binding modes obtained during the first phase. Usually searching algorithms provide more than one solution, i.e. more than one pose (or binding mode) for a particular protein-ligand complex that must be evaluated. At the same time, in the context of VS simulations, all the ligands must be ranked according to their propensity to bind the protein of interest in order to prioritize only a subset of molecules [27] promoting them to further studies and analysis. This task is performed by the scoring functions.

Generally, the association (noncovalent) between a protein (P) and a ligand (L) to form the complex (PL),



is described by the equilibrium (or association) constant  $K_a$  or its reciprocal dissociation constant  $K_d$ :

$$K_a = \frac{[PL]}{[P][L]} = \frac{1}{K_d} \quad (2)$$

The two constants are related to the standard free energy of binding as follows:

$$\Delta G_b^\circ = -k_B T \ln(K_a C^\circ) = k_B T \ln\left(\frac{K_d}{C^\circ}\right) \quad (3)$$

where  $k_B$  is the Boltzmann constant and  $C^\circ$  is the standard concentration. As it is common to all entropy-related quantities, sampling remains a major problem and together

with other approximations, the estimation of the free energy of binding is often a complex task. From a statistical mechanics standpoint, equation (3) translates to [28]:

$$\Delta G_b^\circ = -k_B T \ln \left( \frac{Z_{PL}}{Z_P Z_L} C^\circ \right) \quad (4)$$

where  $Z_{PL}$ ,  $Z_P$ , and  $Z_L$  are the configurational integrals of the bound complex, the protein, and the ligand, respectively. For each  $i$ -th species, the configurational integral  $Z_i$  can be expressed as [28]:

$$Z_i = \int e^{-(U(\mathbf{x}_i) + W(\mathbf{x}_i))/k_B T} d\mathbf{x}_i \quad (5)$$

where  $U(\mathbf{x}_i)$  and  $W(\mathbf{x}_i)$  are the potential energy and the solvation free energy of the  $i$ -th component in the configuration  $\mathbf{x}$ . At this point, it is evident that both solvation and desolvation effects (and the inherent entropy quantities) caused by the binding event must be carefully taken into account to evaluate the configurational integrals and the relative free energy of binding.

Unfortunately, as already mentioned, molecular docking procedures aim at being fast and the evaluation of these quantities would require more expensive computational procedures. Therefore, a number of different strategies have been implemented in order to address this daunting task and to reduce the computational burden, mostly introducing approximations in solvent treatment (force-field based scoring functions) or by adopting phenomenological descriptions (empirical and knowledge-based scoring functions). Nowadays, there are a plethora of different scoring functions and the user can decide which one (or which combination of them) to use for a specific task (for a comprehensive list, see [29]).



### 1.1.3. Combining Docking and Molecular Dynamics Simulations

To sum up, there are no doubts about molecular docking capabilities at identifying crystallographic binding modes and at selecting promising compounds from large databases of molecules when protein flexibility does not play a relevant role. Likewise, its limitations and weakness regarding scoring and sampling problems are clear when dealing with protein-ligand complexes that require large loop rearrangements rather than small local movements. In principle, all these problems can be overcome by molecular dynamics (MD) simulations and related methods [30].

The initial idea was that of supporting molecular docking with MD simulations with the purpose of giving insights into dynamic aspects of the binding event. The serial use of the two approaches, in fact, aims at solving the flexibility problem or, at least, making docking results less subjected to the protein model(s) used to evaluate ligands binding modes and affinities.

Clearly, MD simulations can be performed prior or after molecular docking and the order in which they take place is relative to the issue that one wants to address. In the framework of ensemble-based VS, for example, the ensemble of protein conformations is generated running MD simulations before doing docking (Figure 1). Doing so, protein flexibility is taken into account as a series of different conformations (that are “snapshots”, frames of an MD trajectory) against which a series of parallel VS is run. It can be cited the “relaxed complex scheme” developed by Lin *et al.* [31] and improved by Amaro *et al.* [32] according to which a long MD simulation of the protein of interest in the apo-form is run in order to extract, at regular intervals, different protein conformations. Subsequently, parallel VS are run against each of the extracted MD frames. However, it is not trivial to choose the appropriate protein conformations to build the ensemble and, as it has been shown, the use of too many conformations could be detrimental rather than useful for VS results [24]. In this respect, with the aim of selecting from the MD simulations only that structures that are clearly dissimilar so as to reduce redundancy and improve VS efficiency, cluster analysis approaches can be very helpful [33–35].

On the other hand, MD simulations can be started from docking outcomes with the aim of validating or refining docking results with higher-level theories or sampling approaches (Figure 2). In this case, MD simulations may be useful to address two different problems, one related to posing phase and the other related to scoring phase of docking. In respect to

the former, MD simulations can reveal unstable binding modes, helping in filtering out physically unreliable docking solutions or even identifying new ones [36–38]. In particular, MD simulations allow accounting for protein flexibility and induced-fit effects, although only in a second phase. In respect to the latter, instead, MD simulations can be used to refine energetics estimated by the simple scoring functions used by docking algorithms, with re-scoring purposes. A number of re-scoring approaches are currently available, based on different theoretical methods [39], but they approach a more rigorous description of binding computing energies as ensemble averages. Re-scoring schemes range from partially empirical methods such as Linear Interaction Energy (LIE [40]) through to authentic free-energy approaches such as Free-Energy Perturbation (FEP [41]) or Thermodynamic Integration (TI [42]). It has also to be cited the Molecular Mechanics-Poisson-Boltzmann Surface Area and Molecular Mechanics-Generalized Born Surface Area (MM-PB/GB SA) re-scoring method, maybe the most popular among them as it provides a good balance between reasonable accuracy and computational costs [40].

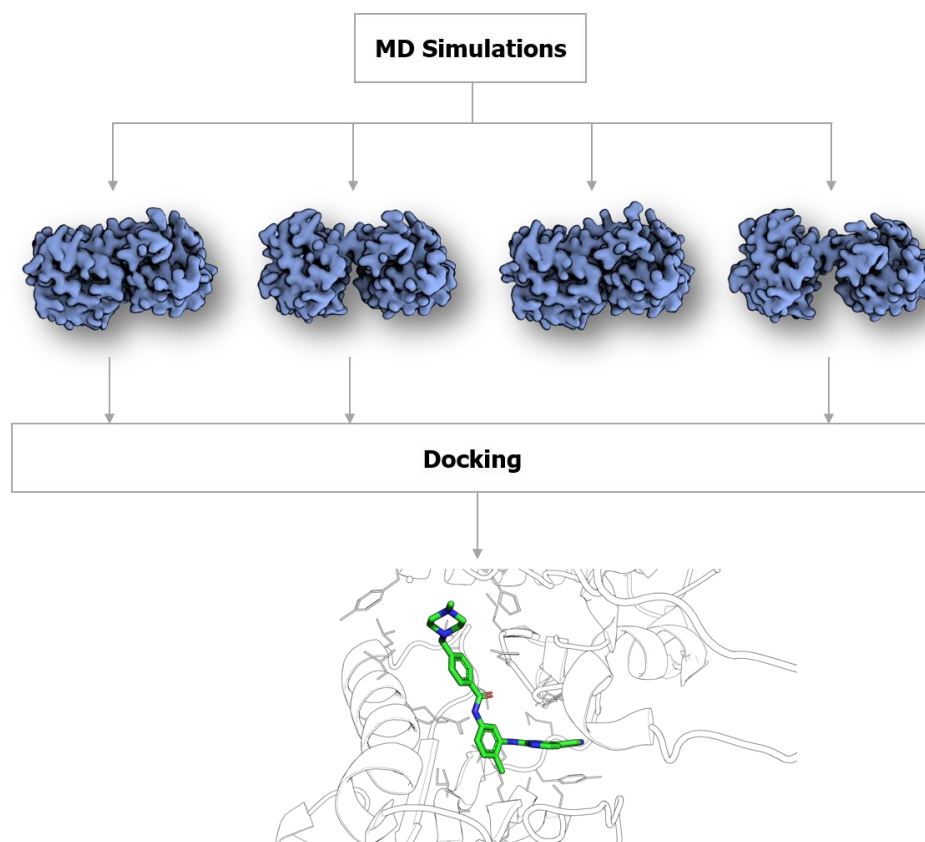


Figure 1 – MD simulations are employed for generating a conformational ensemble of protein structures. A number of snapshots are selected by cluster analysis and then docking is performed against these multiple rigid receptor conformations to evaluate protein flexibility.

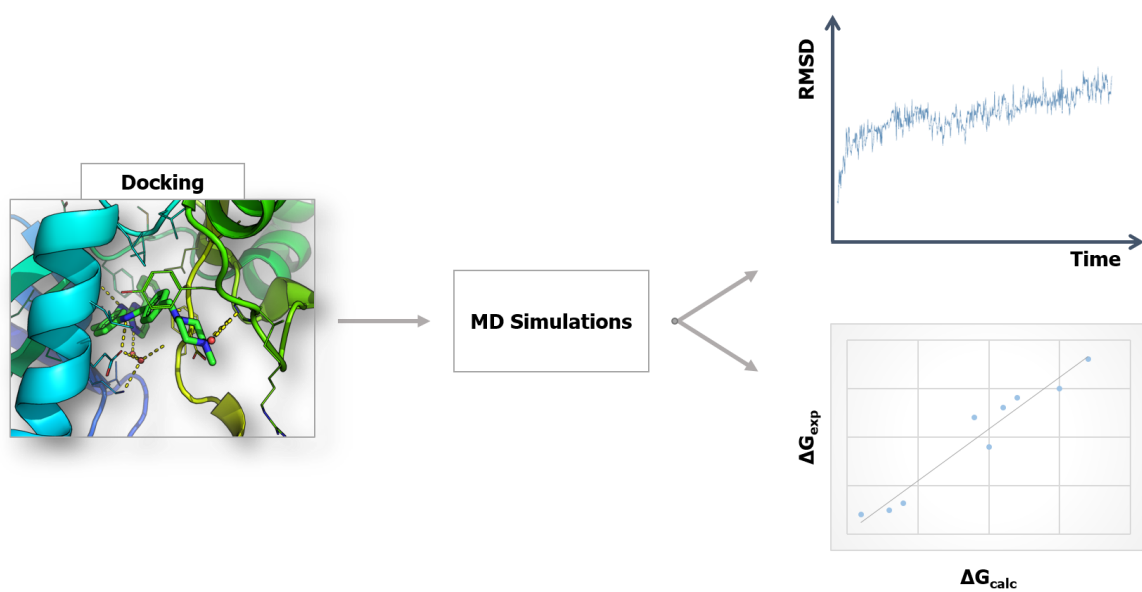


Figure 2 – MD simulations run after docking with the purpose of refining and rescoreing the obtained binding modes.

## 1.2. Towards Dynamic Docking

As it is now clear, standard molecular docking procedures handle the problem of protein flexibility only partially or, when combined to MD simulations, in two different moments, in a way that protein and ligand degrees of freedom are not treated simultaneously, as it should be. At the same time, they lack the required accuracy to obtain a reliable estimation of binding free energy and cannot be used to estimate kinetics quantities. Hence, a paradigm shift from a static perspective to a dynamic approach of the binding process is now inevitable [43,44].

With the advent of faster (mainly Graphics Processing Unit (GPU)-based clusters [30]) and specific-developed (such as ANTON, design and built by D.E. Shaw Research [31]) hardware architectures along with the improvement of computation algorithms, classical MD simulations are gaining everyday more importance in drug design and development programs. MD simulations are inherently able to deal with the structural flexibility of drug-target systems at a fully atomistic description [30]. Moreover, on the condition that simulations are long enough to cover the entire drug-binding process (i.e., from the drug fully solvated in water to the final bound state), thermodynamics and kinetics quantities can be evaluated straightforwardly [30,43]. Nowadays, it is possible to reach and sometimes exceed milliseconds-long MD simulations [45,46]. However, in order to evaluate thermodynamics observables and kinetic quantities, the event of interest (i.e. the binding event, in this case) has to be observed multiple times along the simulation so as to thoroughly sample the configurational space and collect adequate statistics. This usually requires multi-milliseconds simulations making the whole process computationally demanding, even when studying just a single molecule with the help of specialized hardware. In other words, there is a long road ahead before “brute force” plain MD simulations will be routinely used by industry in place of standard docking techniques at the beginning of the long drug discovery process when novel compounds have to be identified. The same applies when a series of compounds in the hit-to-lead optimization have to be evaluated. To overcome these limitations, during the last years, a number of different enhanced sampling methods [47] have been introduced. These approaches, by applying biasing forces or altering the potential energy function allow researchers to accelerate MD simulations increasing the probability of observing the event of interest, such as ligand-protein (un)binding. Thanks to these approaches and to the ever-increasing computational power, MD simulations are going to

replace the old-fashioned view of static molecular docking with a novel concept commonly referred as “dynamic docking” [43,44].

The main difference in respect to the traditional docking techniques is that dynamic docking simulations aim at characterizing the protein-ligand binding process at a fully dynamic level. However, similarly to traditional docking methods, there are two issues that need to be addressed: the first is related to the ability to generate binding modes, which strongly depends on the sampling strategy used (classical unbiased MD simulations vs enhanced sampling methods); the second is related to the evaluation of the identified binding modes that can be done by estimating the free energy of binding or by relying on statistical approaches.

### **1.2.1. Unbiased-MD Approaches**

As previously introduced, the improvement in computer architectures (i.e. GPUs, specialized hardware or distributed computing networks such as GPUGRID.net) allowed several groups reaching and breaking the millisecond barrier in MD simulations timescales. Therefore, what ten years ago was no more than a hope, that is simulating a spontaneous and unrestricted drug-binding event without applying any bias to the system, has now become possible. Unbiased-MD approaches are all that approaches in which the dynamics is not altered through the use of external forces. Even though some workarounds have been proposed to reduce the required computational none of them relies on enhanced sampling methods.

#### **1.2.1.1. Long Plain MD Simulations**

The first example of a spontaneous drug binding events was reported by Shan et al in 2011 [48]. The authors randomly placed two Src kinase inhibitors within a simulation box together with their target protein and let them freely diffuse to their binding site. Not only several spontaneous binding events forming complexes nearly identical to those resolved by X-ray crystallography were recorded, but they also identified a previously unknown allosteric pocket, underlining the potential of these simulations as a suitable tool in standard drug discovery programs. The same trail was trodden by Buch and co-workers in order to completely reconstruct the benzamidine to trypsin binding process in terms of the pathway and related energetics, giving insights into the mechanism of association of a drug to its target without neglecting the intermediate states [49]. Similar works were then conducted for membrane receptors (the  $\beta$ -adrenergic GPCRs coupled to agonist and antagonist small

molecules [50] and spontaneous binding of tiotropium and acetylcholine to M2/M3 muscarinic receptors [51]) and, recently, for the purine nucleoside phosphorylase enzyme and its picomolar inhibitor, the transition state analog DADMe-immucillin-H [52]. Similarly, it has been shown the feasibility of applying unbiased MD simulations in the context of fragment-to-lead development [53], where as well as reproducing all the crystallographic poses of carboxy-thiophene fragments present in the X-ray structure of AmpC  $\beta$ -lactamase, trajectory analysis was able to discriminate distinct binding modes from both a thermodynamic and a kinetic standpoint [54]. Hence, this dynamical approach to docking shed light on ligands route to binding, helping to characterize the major energetic barriers along it and the factors that may be of influence on them such as transient interactions, dehydration, ligand geometry, and so on, many of which mistreated by traditional docking techniques. As a matter of facts, dynamic docking can also be used to reveal binding sites and poses of known binders without having previous knowledge of them. As reported by Dror and co-workers, who studied the M2 muscarinic acetylcholine receptor along with a number of experimentally identified allosteric modulators for which no crystal structure was available. They were able to identify an extracellular-facing vestibule to whom several modulators can bind, observing that the binding driving force was a set of cation- $\pi$  interactions rather than the previously proposed aromatic-aromatic contacts. It is worth noting that their results were validated by mutagenesis analysis and radioligand binding experiments, confirming the predictive power of this kind of simulations [55].

Taken as a whole, brute force unbiased-MD simulations set a striking landmark in the development of dynamic docking but also pointed out their main limitations. The binding of a small molecule to its binding pocket, in fact, can still be considered a rare event especially when considering ligands that have unfavorable on-rates (i.e. low  $k_{on}$ ). To this end, the easiest and most popular approach to overcome this hurdle was to place in the simulation box more than one ligand molecule, so as to increase the probability of observing a binding event. Although this strategy may be of some help, in order to sample thoroughly the configurational space and collect enough statistics, multiple long trajectories are still needed. Therefore, it is not surprising that most of the above-reported applications took advantage of specialized hardware or massive computing architectures to carry out the simulations. That's the reason why, in the context of unbiased simulations, a number of solutions have been proposed to bring dynamic docking to a more practical and effective usage. These strategies include adaptive sampling methods, supervised approaches, and multiscale modeling.

### 1.2.1.2. Adaptive Sampling Methods

Adaptive sampling methods entail a class of unbiased-MD approaches relying on the framework of Markov State Model (MSM) [56,57]. They aim at identifying under-sampled states along the binding process in order to run new simulations from these points (a process usually termed “seeding”) so as to lower the statistical error related to thermodynamic quantities. In this way, few ultra-long simulations are replaced by many short ones, allowing one to save valuable computational time. In the earliest implementations, human intervention was essential to manually select from where restarting a simulation [58], but in 2014, Doerr and De Fabritiis, presented an automated protocol capable of reconstructing the binding process for the trypsin-benzamidine system in a totally unsupervised fashion [59]. The automated learning method showed to be able to achieve a converged binding affinity one order of magnitude faster than classical sampling. They implemented their strategy in a freely available environment (known as High Throughput-MD, HTMD [60] that is also amenable to replace MSM with other adaptive sampling algorithms) and used it to elucidate the cooperative recognition mechanism of ionic cofactors and substrate in the myo-inositol monophosphatase enzyme [61], and the binding process of the lipid inhibitor ML056 to sphingosine-1-phosphate receptor [62]. The latter is an informative example since described the diffusion of the lipophilic ligand ML056 from the bulk solvent to the membrane bilayer prior to moving to its orthosteric binding site that is not accessible from the extracellular environment like other transmembrane receptors.

A strikingly different approach is the supervised molecular dynamics (SuMD) protocol devised by Sabbadin and Moro [63]. While in adaptive sampling methods seeding is employed to resample under-sampled states, in the SuMD implementation, simulations that are unlikely to conduct to a productive binding event are discarded on-the-fly through the use of a tabu-like algorithm. During short simulations windows, at defined time intervals, the distance between the center of masses of the ligand and the binding site is saved and subsequently fitted in a linear function. If the slope of the resulting line is negative it means that the ligand is moving toward its binding site, and the simulation is evolved. Otherwise, the current simulation is interrupted and a new run is restarted from the latest saved coordinates through velocity reassignment. Finally, at a user-defined distance value, the supervision algorithm is switched off and continuous dynamics is restored. The method has been tested against a number of complexes of both globular and membrane proteins [64,65] and showed to be effective in reproducing ligand-receptor complexes in the range of

nanoseconds. On the other hand, its limitation is the need of an a priori knowledge of the binding site location, which, at least in principle, precludes the possibility to discover novel and unexplored binding sites like in brute force dynamic docking simulations. Moreover, while the protocol has proven valuable in identifying relevant metastable states along the recognition process, because of the discontinuous nature of the trajectories, the binding route followed by the ligand is not necessarily close to the minimum free energy pathway.

Similarly, multiscale approaches aim at reducing the time spent by a ligand diffusing in the solvent probing its binding site which is not a relevant part of the association pathway. The idea behind that is to limit the full atomistic resolution and computationally expensive MD simulation only to those regions that are close to the binding site. Very recently, Zeller *et al.* introduced a multiscale approach to Dynamic Docking that also allows the evaluation of kinetics of the binding using as test case two H1N1 neuroaminidase inhibitors, oseltamivir and zanamivir [66]. Their implementation made use of Brownian Dynamics (BD) [67] when the distance between the ligand and the binding site was more than properly defined value. Therefore, in this region, ligands and protein are treated as rigid bodies that undergo translational and rotational diffusion in an implicit continuum solvent. When the ligand reaches the so-called encounter surface, BD switches to all-atom MD and the solvent is treated explicitly. They succeeded in reproducing X-ray crystal structures but, as they stated, this approach could be useful only for those cases where binding pathways are diffusion controlled, having the same usual simulation limitations when the binding involves large conformational changes.



Author (Year)	Complex	Multiple Ligands	No. of Runs	Aggregate Time	Productive Runs <sup>1</sup>	Time to Binding
<b>Brute Force MD</b>						
Shan <i>et al.</i> (2011)	PP1 / Src kinase	y	7	115 $\mu$ s	3	15.1 / 1.9 / 0.6 $\mu$ s
	Dasatinib / Src kinase	y	4	35 $\mu$ s	1	2.3 $\mu$ s
Buch <i>et al.</i> (2011)	Benzamidine / Trypsine	n	495	49.5 $\mu$ s	187	15-90 ns
Dror <i>et al.</i> (2011)	Dihydroalprenolol / $\beta_2$ AR	y	40	111.8 $\mu$ s	5	NA
	Alprenolol / $\beta_2$ AR	y	10	14 $\mu$ s	1	NA
	Propranolol / $\beta_2$ AR	y	21	35.7 $\mu$ s	0	-
	Isoprotenerol / $\beta_2$ AR	y	1	15.0 $\mu$ s	0	-
	Dihydroalprenolol / $\beta_1$ AR	y	10	55.5 $\mu$ s	2	NA
Kruse <i>et al.</i> (2012)	ACh / M3 R	y	1	25 $\mu$ s	1	9.5 $\mu$ s
	Tiotropium / M3 R	y	3	18 $\mu$ s	0	-
	Tiotropium / M2 R	y	3	16.2 $\mu$ s	0	-
Decherchi <i>et al.</i> (2015)	DADMe-immucilin-H / PNP	y	14	7 $\mu$ s	3	340 ns
<b>Discontinuous approaches</b>						
Sabbadin <i>et al.</i> (2014)	ZM241385 / hA <sub>2A</sub>	n	3	-	1	59 ns
	T4G / hA <sub>2A</sub>	n	3	-	1	62 ns
	T4E / hA <sub>2A</sub>	n	3	-	1	105 ns
	Caffeine / hA <sub>2A</sub>	n	3	-	1	15.2 ns
Cuzzolin <i>et al.</i> (2016)	Ellagic Acid / CK2	n	3	-	0	-
	SAPS / GSTP1-1	n	3	-	2	27 - 19 ns
	Benzen-1,2-diol / PRDX5	n	3	-	3	17.4 / 31.2 / 18 ns
	(S)-naproxen / HSA	n	3	-	0	-
	(S)-fluoxetine / LeuT	n	3	-	0	-
	NECA / hA <sub>2A</sub>	n	3	-	0	-
Zeller <i>et al.</i> (2017)	Oseltamivir / neuraminidase	n	676	50.0 $\mu$ s	~20	NA
	Zanamivir / neuraminidase	n	606	35.7 $\mu$ s	~20	NA

<sup>1</sup> “Productive” refers to simulations that reproduced the crystallographic pose within a given RMSD threshold (usually lower than crystal resolution).

Table 1. Comparative time scales of brute-force MD simulations versus discontinuous approaches as reported in retrospective studies. Owing to the inherent difficulties in comparing timescales of several short trajectories, works relying on adaptive sampling methods have not been included in this table.

### 1.2.2. Enhanced Sampling Methods in Molecular Dynamics

As shown in the previous paragraph, even though it is possible to record and sample a binding event using classical MD, these simulations are often too demanding in terms of computational cost to be routinely used in the drug discovery and development process. Therefore, all the works that have been done seem to be more proofs of concept rather than usable and practical tools. Most of the times, because of the large number of the degrees of freedom that are needed to be explored and the time spent by the ligand in the bulk solvent prior to finding its way to the binding site(s), exploiting these methodologies is not accessible with the currently available computational resources. As is often the case with molecular dynamics simulations, the system, eventually trapped in a local minimum, is not able to overcome the energy barriers and follow the path that separates the absolute minimum corresponding to the complex bound state. At the same time, higher energy regions are not well sampled so as to make hard to evaluate free energy differences.

For these reasons, a number of methods, generally referred as *enhanced sampling methods*, have been introduced in order to overcome the energy barriers and improve the sampling of the configurational space by using different and, often, elaborate methodologies but always relying on a full-atomistic description of the event of interest. For the sake of clarity, they can be divided into two categories [30,47]. A first one comprises all the methods that enhance the sampling among one of few user-predefined collective variables (CV) which are functions of the atomic coordinates, such as metadynamics (MetaD[68]), steered MD [69], and umbrella sampling (US, [70]). A second one, in which the configurational space is explored at different temperature values or as a function of the potential energy. Replica exchange MD (REMD [71]), accelerated MD (aMD [72]), potential scaled MD [73], and multicanonical MD (McMD [74]) are examples of methodologies that fall among this category.

Strictly speaking, the strength of the former methodologies is a powerful but complex and delicate tool to handle. Seeing that they rely on the user choice of one or more CVs (see paragraph 1.3.1., On the Choice of the “Right” CV(s)), they give results that have to be carefully evaluated. Moreover, if some important degrees of freedom for the process of interest are not taken into account (mainly because a “wrong” set of CVs is chosen), the calculated free energy relative to the chosen path can have no physical meaning [47].

Tempering methods, instead, are free from this problematic, as they act on all the degrees of freedom of the system at the same time.

Focusing on the results that have been achieved, the following are some of the cases in which enhanced sampling methodologies have been successfully exploited in the dynamic docking contest.

#### **1.2.2.1. Methods Relying on CV(s)**

One of the earliest enhanced sampling methods used to characterize the ligand binding process is the metadynamics (MetaD) method introduced by Laio and Parrinello in 2002 [68]. In MetaD, a history-dependent repulsive potential acting on a few CVs is added to the underlying dynamics, discouraging the system from exploring previously visited regions of the CV space (Figure 3). In 2005, Gervasio *et al.* conducted the first application of MetaD to ligand binding [75]. To the best of our knowledge, this is the first example of dynamic docking. The authors used MetaD to successfully reproduce the binding mode and the experimental binding free energy for four complexes:

- $\beta$ -trypsin/benzamidine;
- $\beta$ -trypsin/chlorobenzamidine;
- Immunoglobulin McPC-603/phosphocoline;
- Cyclin-dependent kinase 2 (CDK2)/staurosporine.

A few years later, Provasi *et al.* reported another successful example of the utility of MetaD for simulating protein-ligand binding events. Here, they investigated the binding pathway of the nonselective antagonist naloxone to the alkaloid binding pocket of a delta opioid receptor [76]. Notably, the authors accurately assessed the association constant from the free-energy profile reconstructed through MetaD. This was made possible by efficiently sampling the ligand in the bulk region, and allowing multiple (un)binding events, which are the key to accurately determining the binding free energy. To do so, they confined the unbound state of the ligand using a conically-shaped restraint whose contribution to the association constant can be taken into account analytically. Similarly, in 2013, Limongelli *et al.* developed the so-called funnel metadynamics, in which the previously described confinement is replaced by a funnel-shaped restraint, further reducing the space to explore for the ligand in the bulk [77]. Despite the successful results in predicting the mechanism of ligand binding, MetaD suffers the same drawbacks of any other CV-based method, i.e., the need to choose an optimal set of CVs [78]. Moreover, the simulation time increases

exponentially with the number of CVs, and the MetaD performance rapidly deteriorates. This makes it difficult to accurately simulate systems characterized by a high degree of complexity. In 2007, to overcome these difficulties, Laio *et al.* developed a new method, bias-exchange metadynamics (BS-MetaD), which combines concepts of MetaD and replica exchange (see paragraph 1.2.1.2., Methods Relying on Tempering) [79]. In particular, multiple MetaD simulations are performed in parallel and exchanged at fixed time intervals. Each replica is biased with a time-dependent potential acting on a different CV, thus alleviating the problem of CV selection. In 2009, Pietrucci *et al.* used the BS-MetaD technique to successfully describe the binding mechanism of a small peptide to the HIV-1 protease [80]. Even though the authors accurately computed the free energy associated with ligand binding and unbinding as a function of 7 CVs, almost 2  $\mu$ s of simulation were required to converge the free energy. Despite this, they managed to characterize the kinetics of the (un)binding process using a discrete-states kinetic model, including the relevant metastable states along the recognition pathway. Another variant of MetaD, reconnaissance metadynamics (RMD [81]), provides a valid alternative to bias-exchange in considering a larger set of CVs. In particular, RMD is a machine-learning approach where the algorithm tunes the applied bias using data obtained from short MD simulations. Compared to conventional MetaD, this procedure relieves the user of the *a priori* selection of a small number of CVs, and thus provides a way to efficiently explore previously uncharacterized mechanisms. In 2012, Soderhjelm *et al.* applied RMD to identify and score protein-ligand binding poses of the well-known trypsin/benzamidine system [81].

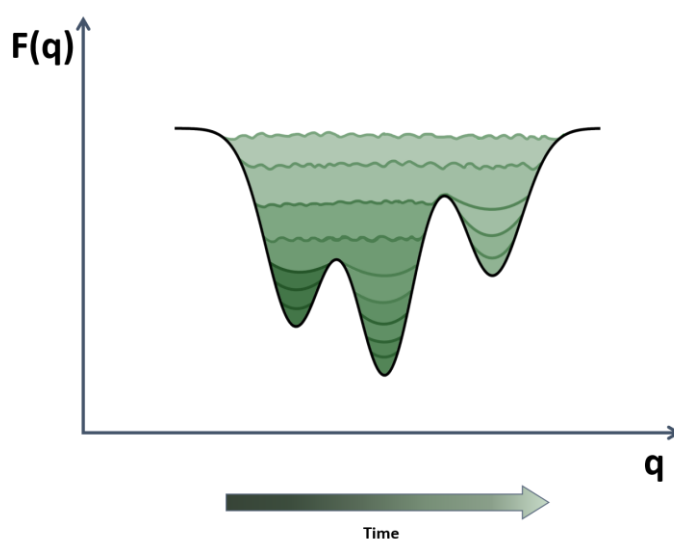


Figure 3 - In MetaD, a bias is applied to a defined CV ( $q$ ) in order to fill the underlying free energy ( $F(q)$ ) and discouraging the system to sample already sampled states.

### 1.2.2.2. Methods Relying on Tempering

Of the various tempering methods, the REMD method has emerged as one of the most widely used techniques to enhance conformational sampling [82]. In classical REMD, several replicas of the system are simulated independently in parallel, at different temperatures. At regular intervals, exchanges between neighboring pairs of replicas are attempted according to a Metropolis acceptance criterion (Figure 4). Because an efficient exchange requires a significant overlap of the potential energies sampled at adjacent temperatures, a high number of replicas is typically required for the method to be effective. Hamiltonian REMD (H-REMD), in which the different replicas are simulated at the same temperature while the system's force field is modified, provides a valid alternative to temperature REMD. An advantage of H-REMD compared to classical REMD is the possibility of varying only part of the Hamiltonian of the system among the replicas, improving the exchange probabilities [83]. This computational framework was adopted by Luitz *et al.* to obtain the correct binding modes for protein-ligand systems through explicit solvent simulations [84]. In particular, the H-REMD approach was based on softening the ligand-protein non-bonded interactions along the replicas in order to prevent the sampling of irrelevant states before reaching the native binding mode. The method was tested on three different systems: human FKBP protein (FKBP-52) in complex with the high-affinity ligand FK506 and with the lower affinity ligand SB3, the peptide-binding domain of murine MHC class 1 molecule in complex with a viral antigen. aMD is another enhanced sampling technique that does not rely on the a priori definition of CVs. aMD speeds up the configurational space sampling by locally adding a non-negative boost potential to the system's potential energy. The potential energy is added only to those regions of the potential energy that are below a certain threshold energy value, while leaving those above this level unaltered [72]. Recently, Kappel *et al.* used aMD simulations to simulate processes of ligand binding to the M3 muscarinic receptor, a G-protein-coupled receptor (GPCR) [85]. In particular, this work used long-timescale aMD simulations (hundred-nanosecond timescale) to identify the metastable ligand-binding sites of three known molecules: the antagonist tiotropium, the partial agonist arecoline, and the full agonist acetylcholine (ACh). Reweighting procedures recover the canonical distribution, from which the free-energy landscape can be calculated. However, it is still challenging to estimate the exact population for each configuration. This is because the reweighting procedure is subject to a statistical error, especially when longer timescales are simulated. In this work, where aMD simulations

were performed on the 100–1000 ns timescale, the authors focused on identifying metastable ligand-binding sites on the M3 receptor, in agreement with unbiased MD simulations, in a significantly shorter time (about 80 times faster for ACh). In 2007, Kamiya *et al.* provided interesting results by performing McMD simulations to successfully dock the inhibitor tri-N-acetyl-d-glucosamine [86]. McMD is an enhanced sampling method, in which a random walk sampling through the energy space is made possible by the bias applied to the system [74]. In the McMD method, higher energy states and lower energy states have an equal probability of being sampled because different temperature regions defined by the bias are simulated simultaneously. A merit of the McMD method is that the canonical ensemble can be reconstructed at 300 K relatively easily by a reweighting procedure. Recently, Bekker *et al.* also performed long McMD simulations to dock the inhibitor CS3 to cyclin-dependent kinase 2 [87]. To accelerate the reproduction of the native complex, the ligand was restrained in a cylindrical region near the binding pocket. In addition, after having identified the correct binding mode, they accurately predicted the binding free energy by TI in accordance with the experimental data.

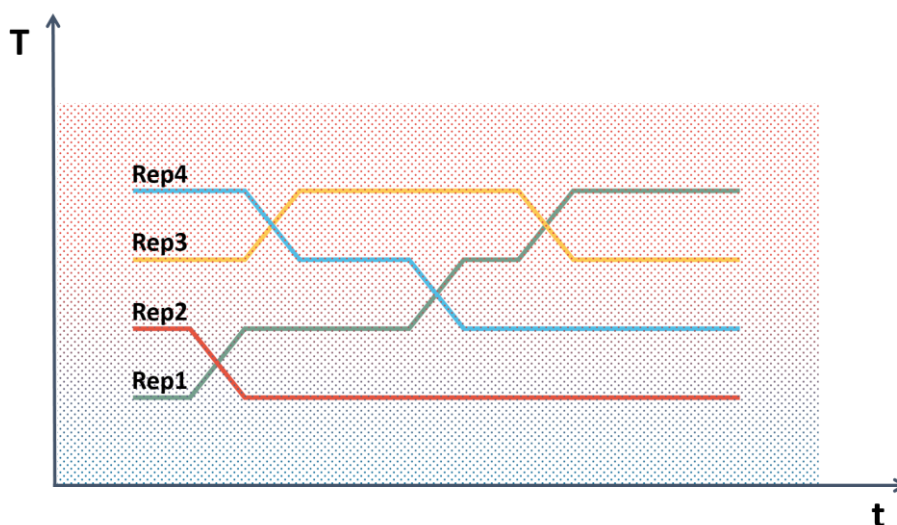


Figure 4 – Among the methods that do not rely on CVs, REMD simulates independently several replicas of the system at different temperature (T) giving them the possibility to exchange coordinates at regular intervals. Background color indicates the Temperature. Rep1-4 are indicated with different colors.

## 1.3. The MD-Binding Approach

### 1.3.1. On the Choice of the “Right” CV(s)

As a matter of facts, the possibility of choosing a suitable CV in a molecular dynamics simulation represents a practicable way to reduce the complexity of the large number of degrees of freedom to a much lower dimensional space [88]. Thus, few simple parameters can be either analyzed or manipulated (biasing or “forcing” the system to follow the reaction coordinates along this path). However, defining a good (or the so-called “right”) CV for a biological process is not a trivial task. The ideal CV has to be representative of the process of interest, and capable of describing it in all its aspects without neglecting other important variables that have to be taken into account for a complete and effective analysis. In other words, the chosen set of CVs may not provide a physical description of the process and of the inherent energetics, obtained evaluating the energetic minima and the barriers crossed by the system along the specified path[47].

Usually, most of the CVs are relative and defined only for a specific issue that one wants to address. Some examples can be the torsion angles for analyzing conformational changes, the center-of-mass (COM) distance between two sets of atoms, likely a molecule and some residues of a binding site in the contest of (un)binding studies, and so on. In principle, it is necessary to possess a good and detailed *a priori* knowledge of the process of interest, and this is not common. Moreover, the set of CVs that have proven to be efficacious in describing a specific process for one system perhaps is not generally transferable to other systems.

At the same time, every chosen CV has to satisfy a number of conditions[47,68,78]:

- It has to be function of the coordinates of the system;
- It has to be defined by a continuous and differentiable mathematical function;
- It has to be capable of describing all the slow events that are relevant the process of interest;
- It has to be able to discriminate among initial, intermediate and final states.

For all these reasons, choosing a set CV(s) is usually a complex task that hides not few difficulties. The choice is mostly related to the experience of the researcher and it is conducted on a trial-and-error procedure.

Still, in the framework of ligand recognition mechanism, is there a common ground to develop a CV that can be extended to every system or, at least, to a vast majority of them? Dealing with the case of DADMe-immucillin-H and PNP, Decherchi *et al.* noticed that the driving force of the binding was a strong electrostatic field, due to ligand and binding site opposite charges. In a broader perspective, it is well-known the role played by the electrostatic interactions in the context of molecular recognition [89–91]. Hence, they started to think of a way to take advantage of this electrostatic field even for that cases in which the binding event is not naturally driven by electrostatics.

Under this assumption, they decided to introduce an artificial electrostatic bias between the binding partners, i.e. the ligand and a set of residues of the binding site, so as to guide the ligand to its target without spending time floating around in the bulk solvent. The idea can be exploited introducing fictitious charges on the two sets of interacting atoms with a functional form of the type [92]:

$$C \sum_{a \in A, b \in B} \frac{Q_a Q_b}{r_{a,b}} d(r_{a,b}, \lambda) \quad (6)$$

where A and B are the two interaction partners, and  $a$  and  $b$  are the atoms comprised in A and B, respectively, that are selected by the user,  $r_{a,b}$  is the distance between each of them,  $C$  is a modulation parameter, and  $d(r_{a,b}, \lambda)$  is a function that modulates the effect of the potential according to the distance  $r$  and a parameter  $\lambda$  (see below). Conventionally, these charges  $Q_i$  have opposite sign and are spread among all the atoms composing each set of atoms in order to make the process as natural as possible without forcing unusual interactions [92,93].

### 1.3.2. An Adaptive Bias for Ligand Binding

The next question to address is related to the implementation of this CV in a standardized procedure. The ligand has to be guided, rather than “pushed”, to its naturally occurring binding mode in a way that the whole procedure is performed as gently as possible avoiding unphysical pathways. With this in mind, Spitaleri *et al.* devised an adaptive behavior of the external bias acting on the ligand introducing a series of conditions that have to be satisfied. They can be summarized as follows:



- The strength of the bias is user-defined but is always kept at a fraction of the intensity of the physical forces naturally felt by the ligand;
- The biasing force is gradually lowered as the process moves forward, which means that the ligand is moving to its binding site;
- Once switched off, the biasing force cannot be turned back on even if the ligand leaves the binding site moving further to the solvent.

Mathematically, this is obtained modulating the coefficient  $C$  of equation (6) during the course of the simulation in a way that the modulus of the additive forces acting on the subset of ligand atoms (e.g.  $a \in A$ ) is kept at a predefined fraction of the modulus of the overall force originated by the gradient of the regular potential energy of the system. Besides, the modulating function  $d(r_{a,b}, \lambda)$  of equation (6) assumes the form of an exponential decay function:

$$d(r_{a,b}, \lambda) = \exp\left(\frac{-r}{\lambda}\right) \quad (7)$$

In order to delineate a way to measure the distance from the binding site, they introduced the definition of a subset of atoms ( $\tilde{B}$ ) of the selected residues ( $B$ ) of the binding site as “switching-off” residues. The choice is left to the user and should be guided either by *a priori* knowledge of the system of interest, or by considerations leading the user to identifying which pocket residues are potentially relevant for ligand binding. As an example, it can be based on NMR studies or on user assumptions, i.e. one of the deepest residues of the binding pocket, for that cases where no data is available.

The switch-off is obtained via a scaling pre-factor  $\gamma$  that modulates the biasing force, calculated via a switching function as follows [92]:

$$\gamma = \frac{1}{1 + \exp(-ss * (dist - th))} \quad (8)$$

where  $ss$  and  $th$  are two parameters and  $dist$  is evaluated as follows [92]:

$$dist = \min_{x \in A} \min_{y \in \tilde{B}} ds(x, y) \quad (9)$$

where  $ds(x, y)$  is the pairwise distance between the two atoms  $x$  and  $y$ . In this way the bias is switched-off when any atom of the ligand  $A$  falls below a predefined distance from any atom belonging to the subset  $\tilde{B}$  of  $B$ .

It is clear that the user, as is the case with all the methodologies relying on CV(s)-biasing and for any SBDD study, should have an idea about the location of the binding site for the system of interest. However, the main advantage in respect to a traditional molecular docking approach is the possibility of depicting the binding process as a whole, from the ligand in the bulk, fully solvated, to the final bound state, without neglecting the eventual metastable states. There is also another point that has to be highlighted: it can be done at a fraction of the time required by long unbiased MD simulations.

The protocol is now available in BiKi Life Sciences [94] developed by BiKi Technologies.

## 2. Tubulin as a test case

### 2.1. Microtubules Structure and Functions

Microtubules (MTs), together with microfilaments and intermediate filaments, are protein polymers composing the cytoskeleton of eukaryotic cells whose main function is maintaining cell shape and structure [95]. They serve as a rail for the internal trafficking of vesicles and organelles driven by kinesin and dynein motor proteins [96] and play a central role in cellular division (mitosis and meiosis), being the major constituents of the mitotic spindle, the cellular apparatus responsible of separating sister chromatids.

MTs are hollow cylindrical polymers built from the lateral association of usually 13 protofilaments [97] made up of  $\alpha/\beta$ -tubulin heterodimers bound in a head-to-tail fashion (see Figure 5 and Figure 6). However, *in vitro* [98], in bacteria [99] or for other reasons (i.e. the interaction with proteins or drugs [100]), it is possible to observe MTs formed by a different number of protofilaments.

In general, the most common assembly is referred as “13-3” [101], where the first number, 13, indicates the protofilaments that are bound together to form the MT, and the second, 3, the vertical offset between tubulin monomers due to the helicity of the turn. In this regard, MTs with different vertical offsets have been observed [102]. This offset between tubulins causes a pseudo-helical arrangement with a lattice discontinuity called the seam (red line in Figure 7). With respect to its dimensions, for the 13-3 assembly, the MT external diameter is about 25 nm whereas the internal about 14 nm. The length varies depending on its polymerization status and can reach up to 50  $\mu\text{m}$  [103].

The order in which tubulin dimers are added to the growing polymer generates a polarity at the ends of the resulting structure, which is an essential feature of microtubules. Conventionally, the end that is capped by  $\beta$ -tubulin is known as “plus end”, and the other, capped by  $\alpha$ -tubulin, “minus end” (see Figure 5). These terms arise from the rate of growth that takes place at the two ends, which is much faster at the plus end than at the other [102] as observed in *in vitro* studies. In the same context, the polarity of the MT is essential to the function of many MAPs (Microtubule-Associated Proteins), which, thanks to this property, can recognize the plus end and alter MT dynamics guiding the growth toward specific cellular locations, especially during cell division [104].

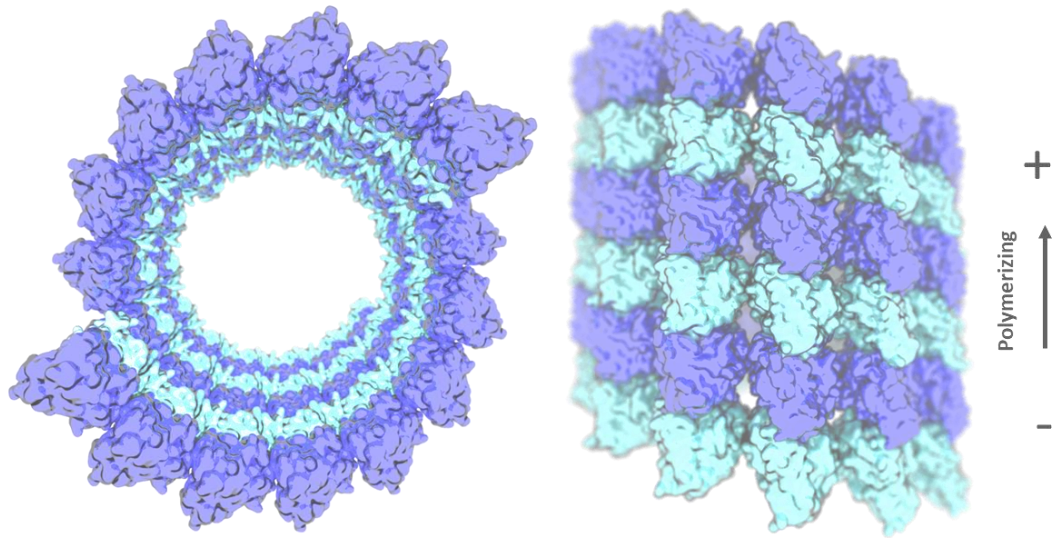


Figure 5 – Structure of a 15 protofilaments MT generated from the cryo-electron microscopy (cryo-EM) in complex with the *Drosophila melanogaster* kinesin-13 KLP10A (PDB ID: 3J2U - 10.8 Å) after the removal of the kinesin protein. From the top view (on the left) it is possible to observe the internal lumen that is characteristic of MTs. The way tubulin dimers are assembled generates a polarity in MT structure (side view, on the right).  $\beta$ -tubulin monomers are depicted in blue surfaces,  $\alpha$ -tubulin in cyan.

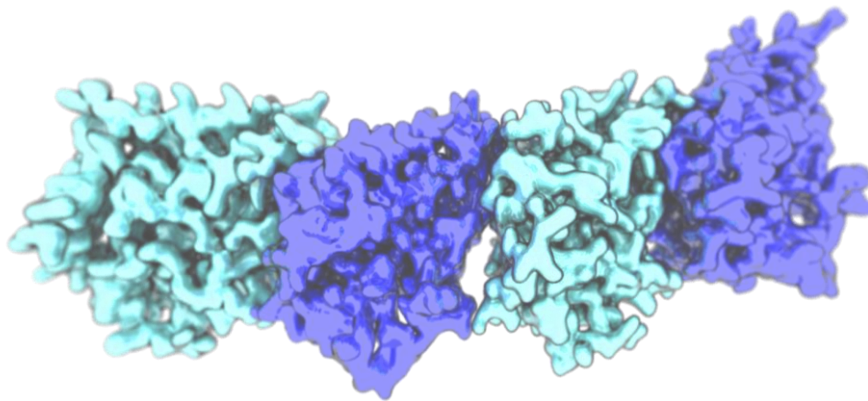


Figure 6 – Tubulin heterodimers.  $\beta$ -tubulin monomers are depicted in blue surfaces,  $\alpha$ -tubulin in cyan.

Inside the cell, MTs are usually anchored through their minus end to the microtubule organizing centers (MTOCs), the major of which in animal cells is the centrosome [105]. MTOCs serve as an initiation site for MTs growth, where the  $\gamma$ -tubulin ring complex ( $\gamma$ TuRC, formed by the association of  $\gamma$ -tubulin monomers together with the gamma complex associated proteins 2-6 (GCP2-6)) specifically interacts with the MTs minus end. In particular, the pseudo-helical-shaped  $\gamma$ -tubulin cap forms strong longitudinal interactions with the  $\alpha$ -tubulin providing a template for the nucleation of the growing polymer [106]. In a broader perspective, nucleation does not always rely on a template and *in vivo* MAPs, and other factors, affect this crucial step [107]. It is worth noting that, *in vitro*, MTs showed the ability to nucleate starting from solutions containing only purified tubulin heterodimers, GTP, and  $Mg^{2+}$  ions [108], despite this spontaneous nucleation is a kinetically restrained process that requires high tubulin concentrations (for a general view of the process, see Figure 7).

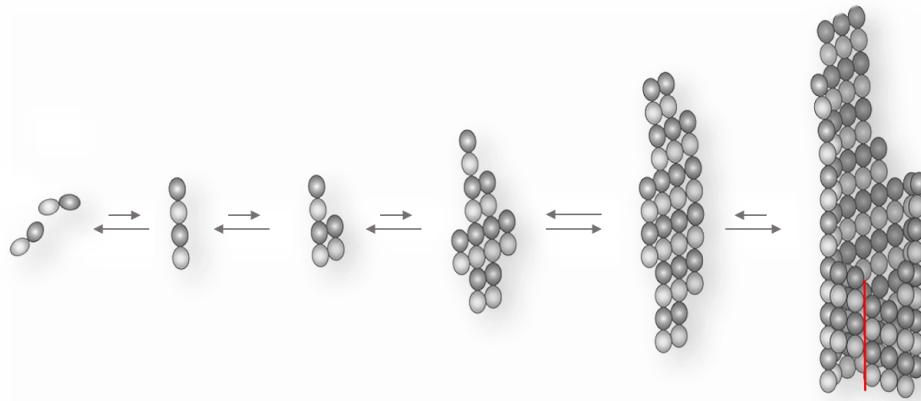


Figure 7 – Spontaneous MTs nucleation and growth. The figure depicts the whole process, starting from a solution of purified tubulin heterodimers with GTP and  $Mg^{2+}$  (on the left) to the complete assembly (on the right). The red line indicates the seam, i.e. the discontinuity in the lattice due to the vertical offset between tubulin monomers.

## 2.2. MTs Dynamic Behavior

As already mentioned, in order to carry out the functions they are responsible for, MTs exploit major conformational changes and go through large variations in their spatial arrangement, likely their foremost characteristic, involving a continuous switching between polymerization and depolymerization phases.

MTs can exert their dynamic behavior through two different mechanisms known as *treadmilling* and *dynamic instability*.

Treadmilling (as shown in Figure 8), in a similar manner as for actin filaments, involves the addition of tubulin heterodimers at the plus ends of MTs, and the simultaneous loss at the minus end. Albeit this process has been observed *in vivo* [109] and may be important in mitosis [110], the resulting biological implications still have to be clarified [111] seeing that, inside the cell, most of the time, MTs' minus end is anchored to the MTOCs. In this respect, dynamic instability [112–114], the mechanism that involves the addition and the loss of tubulin subunits at the same MTs' end, appears to be the most reasonable and acquires a major role in MTs dynamics. The switch from polymerizing to depolymerizing phases is referred as “catastrophe”, the reverse as “rescue” [115].

Dynamic instability (Figure 9) is mainly driven by GTP hydrolysis [116]. Going into details, each tubulin monomer binds one nucleotide; however, there is a significant difference between the two sites. The nucleotide in the  $\alpha$ -tubulin site cannot be exchanged and cannot be hydrolyzed, therefore it has been called N-site (nonexchangeable-site) nucleotide, whereas the one in the  $\beta$ -tubulin is both exchangeable and hydrolyzable, and is referred as E-site (exchangeable-site) nucleotide. Tubulin heterodimers can be incorporated in the growing filaments only when a GTP molecule is present in both sites. After the assembly, only at the level of  $\beta$ -tubulin E-site, GTP is quickly hydrolyzed [117] to GDP, which means that MT lattice has an equal concentration of GTP (at the level of  $\alpha$ -tubulin N-site) and GDP. This condition is necessary for MTs depolymerization, as GDP-bound tubulin is less stable and promotes depolymerization. For instance, in experiments where GTP has been replaced by a slowly hydrolyzable analog (GMPCPP) it has been possible to observe MT polymerization but not the subsequent disassembly [116].

As a matter of facts, the nucleotide at the N-site is responsible for the curvature of the heterodimer: GTP-bound dimers are straighter than GDP-bound ones (for a schematic representation see Figure 8) and this feature can somehow explain why the hydrolysis is linked to the stability of MTs and the related dynamic capabilities [117]. In addition, for reasons that have still to be clarified, in the MT lattice it is possible to find “islands” of GTP-bound  $\beta$ -tubulin that may act as a trigger for switching from the catastrophe to the rescue phase in the context of MTs dynamic instability [117].

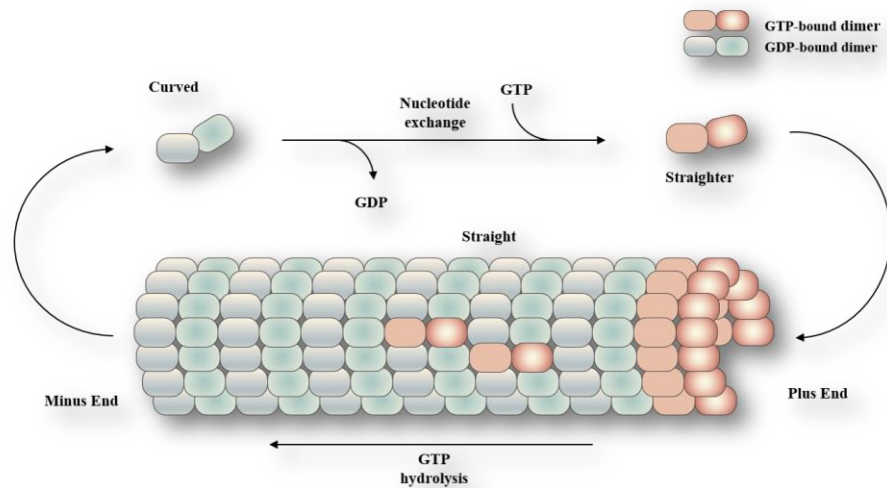


Figure 8 – The influence of nucleotide hydrolysis on tubulin dimer curvature and MTs treadmilling. GTP-bound tubulin islands (in red, at the center of the MT) may act as a trigger for switching among depolymerizing/polymerizing phases.

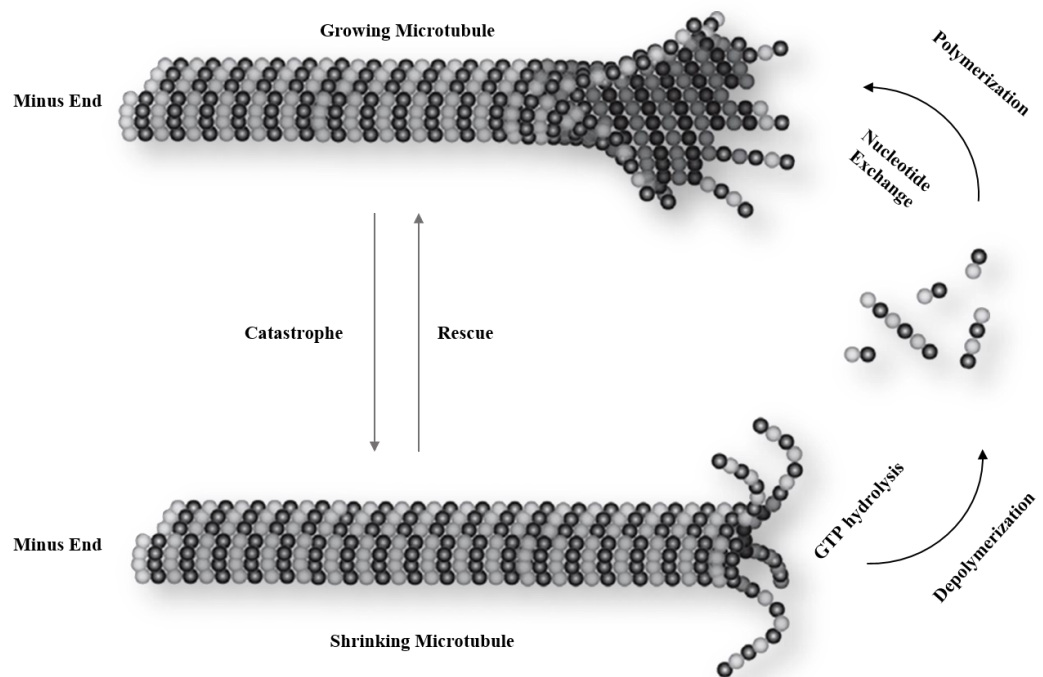


Figure 9 - Microtubule dynamic instability. GTP-bound tubulin assembles at the microtubule plus end. After GTP hydrolysis occurs, the microtubule becomes unstable and depolymerizes at the same end. When GDP is exchanged with GTP in the free tubulin dimers, the cycle can begin again.

## 2.3. MTs bricks: $\alpha$ and $\beta$ -tubulin

As it is now evident, tubulin is a highly dynamic entity that can exist in three major forms: heterodimers, curved oligomers, and assembled in microtubules. For this and other reasons (i.e. the instability of the protein in solution, post-translational modifications and so on) gathering three-dimensional (3D) tubulin crystals is a complex task. The first 3D structure of the tubulin dimer was reported by Nogales *et al.* in 1998 [118] obtained by electron crystallography (EC) of zinc-induced tubulin sheets stabilized by Taxoetere at a resolution of 3.7 Å. The authors were then able to refine their model obtaining a  $\alpha/\beta$ -tubulin structure in complex with Taxol at a resolution of 3.5 Å [119] (Figure 10). The use of the zinc showed to be a good way of incorporating tubulin dimers in two-dimensional sheets [120] that are ideal samples for EC studies. However, the resulting protofilaments, albeit similar to those in microtubules, are arranged in an antiparallel way, which, in principle, can give rise to artifacts. Nevertheless, the same group was able to dock the above-mentioned crystal structure in a 20 Å density map of frozen-hydrated microtubules obtained by cryo-EM [121]. The 15 protofilaments MT showed to be useful for giving insights into tubulin dimer orientation within the MT with respect to the inner lumen, and for identifying lateral and longitudinal interactions between other dimers.

Both  $\alpha$  and  $\beta$ -tubulin have a molecular weight of approximately 50 kDa. Each subunit is composed of ~450 residues, 40% of which are homologous. The authors [118] proposed to divide the structure of each monomer into three functional domains (see Table 2 and Figure 10), although there were no clear differences in the EC density map. The first domain, the N-terminal, constitutes a Rossmann fold [122], typical of those proteins that bind nucleotides, therefore called *nucleotide-binding domain*. The second, a central domain, referred as *drug-binding domain*, which binds Taxol and its derivatives. A final *C-terminal domain*, formed by two helices that within the MT are located on the outer surface, is probably involved in the binding of MAPs and motor proteins. The last C-terminal residues, consisting of intrinsically disordered tails, were not resolved in the model. Most of the differences among different tubulin isotypes (taking human tubulin as an example, there are 6 isotypes of  $\alpha$ -tubulin and 7 of  $\beta$ -tubulin [123]) are located in this region that is also the hotspot for a complex series of post-translational modification involved in the regulation of MT dynamics [124,125]. These differences give rise to what is now referred as “the tubulin code” and more information about it and its implications in various diseases can be found elsewhere [126,127].



Residues	Name	Functions	Secondary Structures
1-205	Nucleotide-binding domain	Binds the nucleotide GTP/GDP	6 parallel $\beta$ -strands (S1-S6) 6 helices (H1-H6)
206-381	Drug-binding domain	Binds Taxol, Colchicine and their derivatives - Contains the “M-loop” involved in lateral contacts between protofilaments	4 $\beta$ -strands (S7-S10) 4 helices (H7-H10)
382-end*	C-Terminal domain	Binds MAPs and motor proteins - Hotspot of post-translational modifications that give rise to the “tubulin code”	2 antiparallel helices (H11-H12)
* Residue number is different for $\alpha$ and $\beta$ -tubulin and may vary among different isoforms.			

Table 2 – Tubulin functional domains

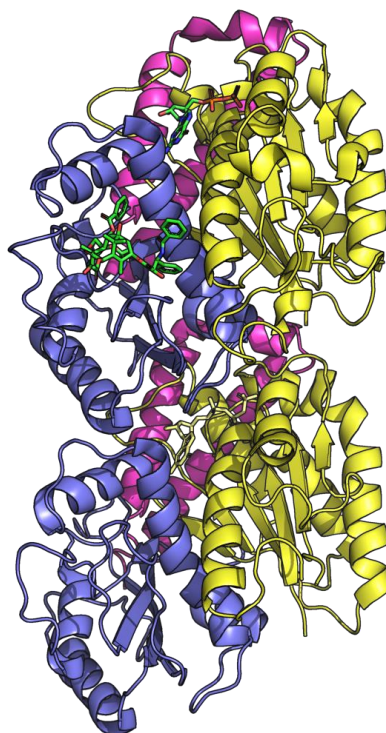


Figure 10 – Tubulin in complex with Taxol (PDB ID: 1JFF – 3.5 Å) and GTP (in the  $\alpha$ -tubulin, down in the picture) and GDP ( $\beta$ -tubulin, up). Functional domains, as reported in [118], are colored differently: N-terminal domain in yellow; drug-binding domain in slate blue; C-terminal domain in light magenta.

A step forward in tubulin structural characterization was represented by the work of Knossow and his group [128]. In 2000, they discovered a new way of stabilizing tubulin for crystallization. Using the stathmin-like domain of the neural protein RB3 (RB3-SLD), a stathmin family protein, they succeeded in obtaining an X-ray structure of two tubulin heterodimers in the “curved” conformation. Stathmin proteins are a family of proteins involved in the regulation of microtubule dynamics. They are able to inhibit tubulin polymerization sequestering two  $\alpha/\beta$ -tubulin heterodimer thanks to their “stathmin-like domain” [129,130], and, from another point of view, they can increase the catastrophe rate of MTs [131].

The T2R structure (T2 stands for two  $\alpha/\beta$ -tubulin heterodimer, R for RB3-SLD) provided a clear model for tubulin sequestration by stathmin family proteins. Moreover, the curvature of the complex is similar to that of the depolymerization products of MTs bound to GDP, suggesting that the binding of the stathmin-like domain does not affect the overall shape of curved tubulin protofilaments [131].

Figure 11 shows the structural differences between the two complexes.

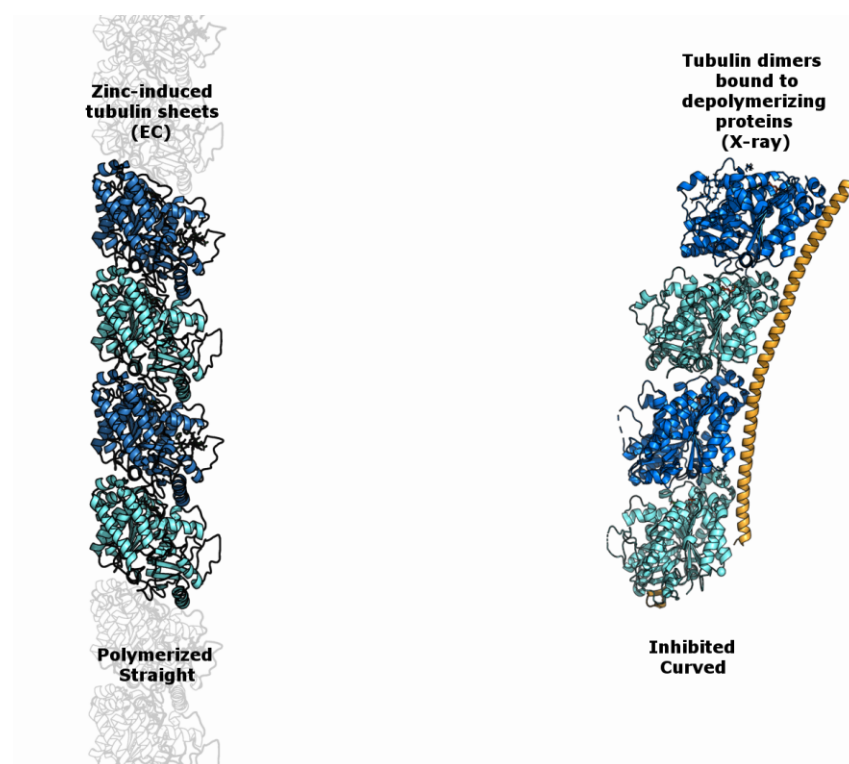


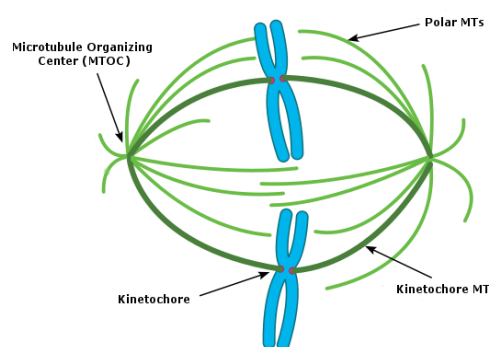
Figure 11 – Structural differences between heterodimers in straight polymerized protofilaments as within microtubules (on the left), and in curved complexes as it happens upon binding of the depolymerizing protein stathmin (on the right).

Their work paved the way for the further studies that brought to the discovery of almost all tubulin binding sites, starting from the colchicine [132] and the vinblastine [133] sites to the recent peloruside/laulimalide [134] and maytansine [135] sites. The same can be said regarding the tubulin tyrosine ligase (TTL) protein that catalyzes the post-translational re-tyrosination of detyrosinated  $\alpha$ -tubulin, whose molecular mechanism has been elucidated in 2013 when Prota *et al.* have solved a T2R-TTL complex [134].

Indeed, most of these structures rely on the assumption that these drugs can bind to both free tubulin and assembled MTs providing only a preliminary and static picture of the complex process that is at the basis of their mechanisms of action.

## 2.4. Tubulin as Target in Cancer Therapy

Among all the different cellular functions in which tubulin is involved, from a pharmaceutical standpoint, mitosis could be considered the most interesting one. Contrarily to the interphase, when MTs turnover (the exchange rate of GDP-bound tubulin dimers with soluble GTP-bound ones) is relatively slow, during mitosis, MTs show a much prominent dynamic behavior [136]. In particular, during prometaphase, MTs attached to the two MTOCs have to grow and shorten quickly and continuously in order to “find” and bind the chromosomes at their kinetochores [137] (Figure 12). If only one single chromosome is not attached by the MTs of the two MTOCs and is not correctly aligned during the metaphase, the cell cannot move to the next phase (anaphase) and, eventually, undergoes apoptosis.



---

Figure 12 – The mitotic spindle [138].

Interfering with MTs dynamic behavior has been proven to be suitable for anticancer therapy [139,140] as cancer cells divide more frequently than normal cells, thus they are more vulnerable to antimitotic compounds (the so-called “mitotic poisons” that induce mitotic arrest and cell death by apoptosis).

In light of the success of the drugs derived from this class of compounds, one could argue that MTs represent the best target for cancer identified so far, and it will continue to be so even when more selective approaches will be developed [136]. As a matter of facts, MTs are “popular” also among plants (i.e. algae or superior plants as the famous *Taxus brevifolia*) and animals (such as sea hares). They both produce toxic molecules that target MTs presumably for self-protective purposes. Besides, most of the antimitotic compounds that have been discovered derive from large-scale screens of natural products. It is possible to cite the Vinca alkaloids, which were isolated more than 40 years ago from the *Catharanthus roseus* [141], or Taxol, isolated from *Taxus brevifolia* in 1971 [142].

## **2.5. Chemical Compounds that Interact with Tubulin**

Going into the details, since the first evidences gathered during the seventies, a large number of compounds that bind either the free tubulin dimer in solution or the formed MTs polymers have been identified. Initially, they were thought to have an influence on MTs mass inside the cell but is now clear that is not the case or, at least, the observed effect is an indirect consequence of their direct influence on MTs dynamics [143].

They show large differences in their chemical scaffolds (Table 1Table 3) and act through distinct mechanisms of action. For this reason, usually these compounds have been classified into two main groups:

- Microtubule-destabilizing agents (MDAs);
- Microtubule-stabilizing agents (MSAs).

MDAs can act in three ways: promoting the depolymerization of the MT; preventing the incorporation of tubulin dimers in the growing protofilaments; precluding that the newly added tubulin dimer adopt a straight conformation. Within this group fall *Vinca* alkaloids (i.e. vinblastine, vincristine), colchicine, combretastatins, maytansine, rhizoxin and others.

On the contrary, MSAs promote polymerization and stabilize the polymer, avoiding the subsequent depolymerization. They have also the ability to induce the formation of the polymer starting from GDP-bound tubulin dimers [144], in contrast to what happens in

normal condition. This second category includes Taxol and its derivatives, the epothilones, discodermolide, laulimalide, peloruside, etc.

Despite their different chemical structure (for an overview, see Table 3) and their different effects, they all bind only to the  $\beta$ -tubulin subunit. At the time of writing, five different binding sites have been identified, which are schematically shown in Figure 13. Usually, when discussing MDAs, one can refer to the intradimer interface as the interface between  $\alpha$  and  $\beta$ -tubulin subunits of the same heterodimer. For example, the colchicine binding site is located in proximity of the intradimer interface. In the same way, the interdimer interface is the one that lies between two adjacent heterodimers. Vinca alkaloids bind at this level and hamper the curved-to-straight transition that is necessary to correctly assemble the MT (more detailed lists can be found elsewhere [145–147]).

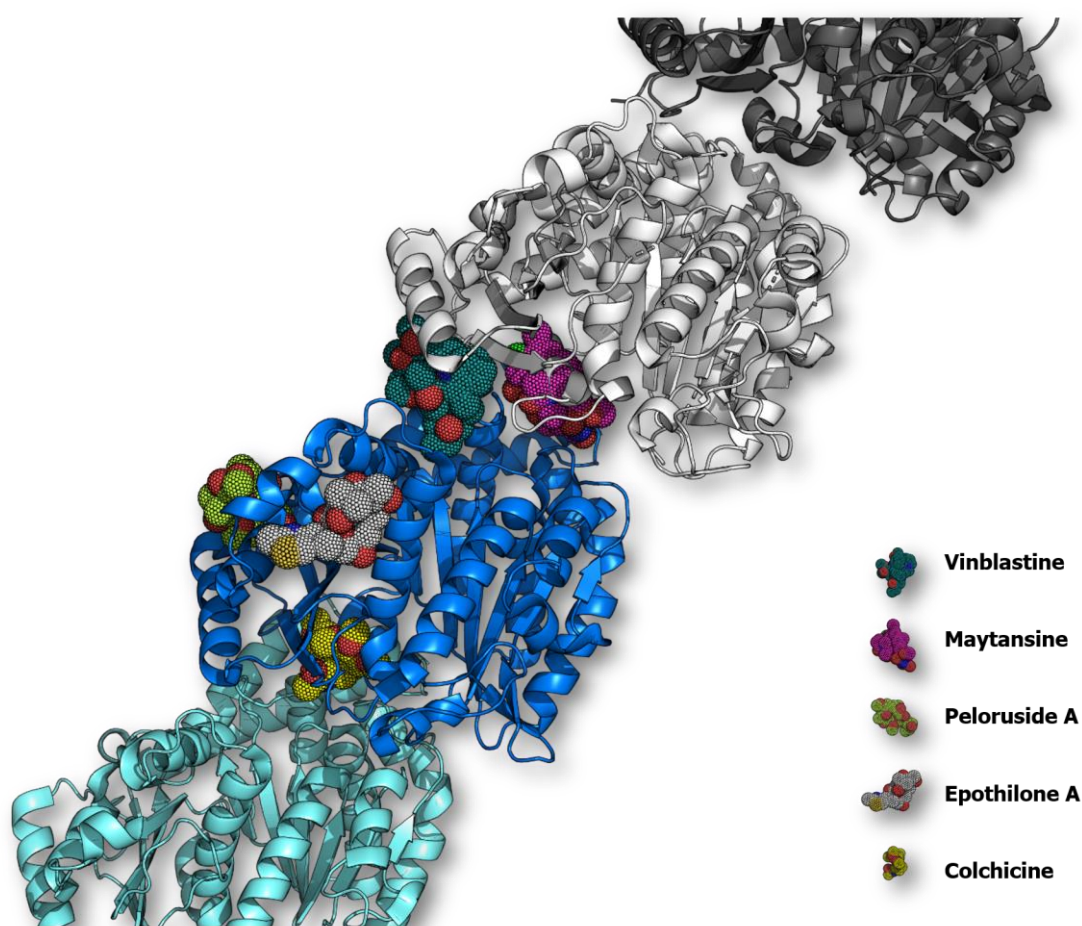


Figure 13 – The five different binding sites in the  $\beta$ -tubulin subunit. Colchicine (yellow dots) binds at the level of the intradimer interface. Vinca alkaloids (dark green dots), such as vinblastine, bind at the interdimer interface, so as maytansine-site ligands (purple dots). MSAs, such as taxane-site ligands (i.e. Epothilone A, white dots) and peloruside A (lime green dots) bind close to the lateral interface between protofilaments.

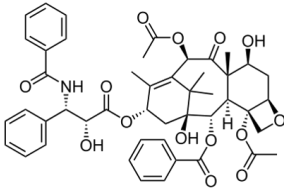
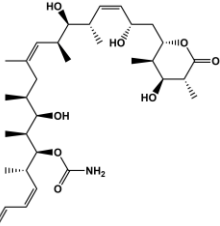
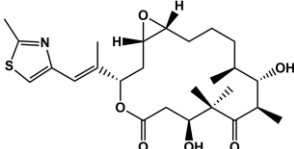
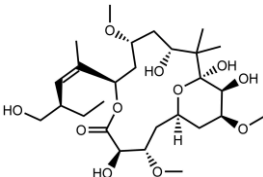
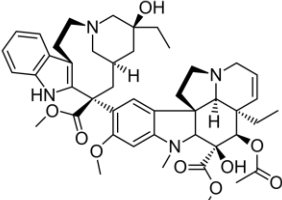
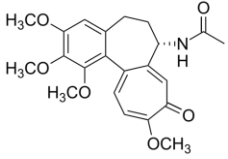
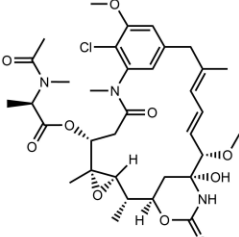
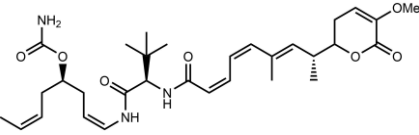
Binding Site	Ligand(s)	
Microtubule-Stabilizing Agents		
Taxane-Site		Taxol
		(+)-Discodermolide
		Epothilone A
Peloruside-Site		Peloruside A
Microtubule-Destabilizing Agents		
Vinca-Site		Vinblastine
Colchicine-Site		Colchicine
Maytansine-Site		Maytansine
		PM060184

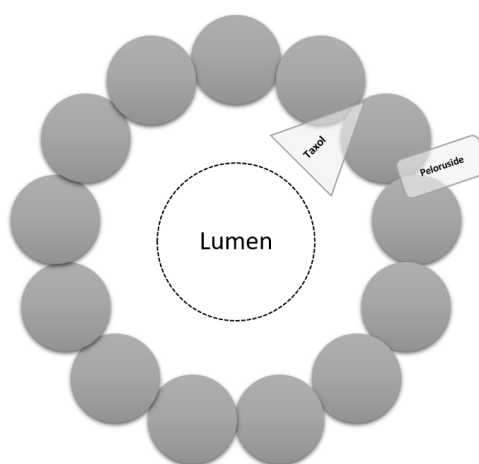
Table 3 – Examples of molecules that interact with tubulin.

From a conceptual perspective, understanding the mechanism of action of the MDAs should be, in principle, easier. The molecules that bind to the maytansine-site exert their activity inhibiting the addition of further tubulin subunits. Similarly, molecules of the vinca domain, as said before, destabilize the MT acting as a wedge.

The same cannot be said for the MSAs. How do they stabilize the MT?

### 2.5.1. The Taxane-Site

Taxol, the first microtubule stabilizing agent described in literature [148], is a complex diterpenoid isolated from the bark of the tree *Taxus brevifolia* [142]. Rao *et al.* first located taxol binding site on  $\beta$ -tubulin in 1992 by direct photoaffinity-labelling [149]. Since then, many structural works confirmed that taxol binds with a 1:1 stoichiometry in the intermediate domain, occupying a site that is closer to the interface between protofilaments facing the lumen of the MT (Figure 14) [119,146,150].



---

Figure 14 – Location of the two MSA sites within the MT shown from the plus-end perspective. Taxol binds on  $\beta$ -tubulin (dark grey spheres) facing the lumen of the MT. Peloruside also binds at the interface between protofilaments, but facing the external part of the MT.

The binding site, located in the drug-binding domain (Figure 10), is composed by the helices H6, H7 (the so-called core helix), part of the helix H1, the  $\beta$ -strand S7 and the loops S9-S10 and S7-H9 (Figure 15). The latter is a very long loop, called M-loop for its characteristic shape, which is directly involved in the lateral interaction between protofilaments [102]. The hypothesis was that the MSAs, pushing the M-loop toward the adjacent protofilament, strengthen this interaction avoiding the depolymerization of the MT.



Unfortunately, the protofilaments in zinc-stabilized sheets are arranged in an alternating, antiparallel fashion. As a consequence, the M-loop interaction that exists within the natural MTs assembly cannot be studied correctly.

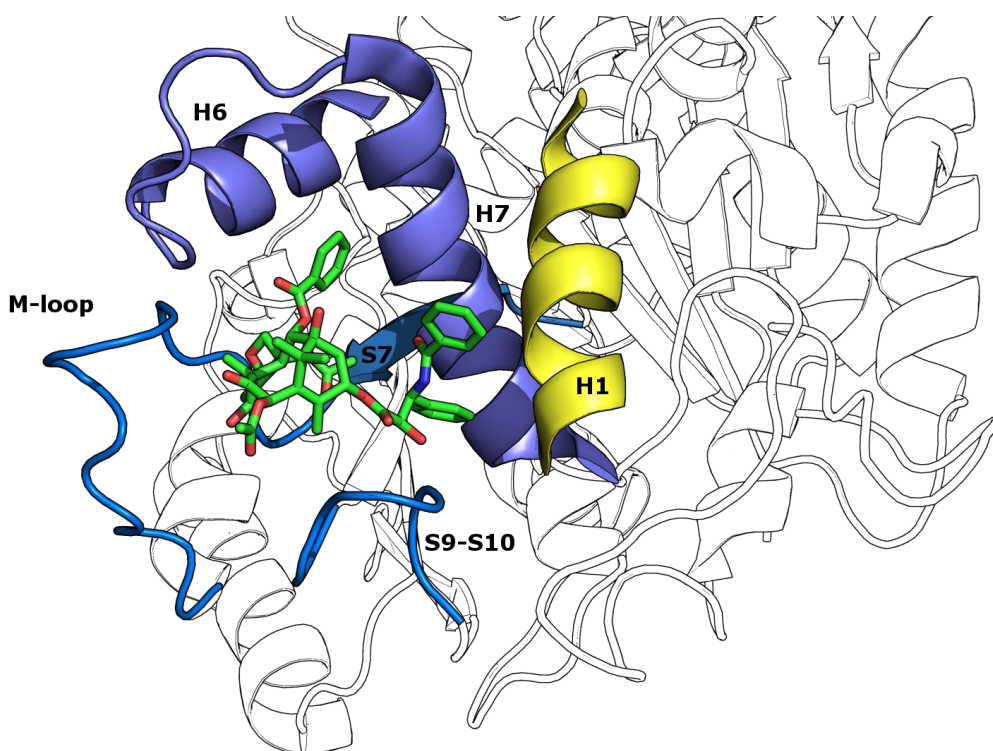


Figure 15 – The tubulin taxane-site when Taxol is bound as in the EC crystal structure (PDB ID: 1JFF)

Despite the early works based their assumptions on these “artificial” zinc-induced tubulin sheets, this theory has been confirmed by the recent x-ray crystallographic studies conducted by Prota *et al.* [135] on others MSAs (Epothilone A and the covalent binder zampanolide). Moreover, in 2014, Nogales and co-workers were able to examine the effects of taxol binding on MT dynamic behavior using high-res cryo-EM structures (4.7-5.6 Å). They noticed that Taxol is able to counteract the effect of GTP hydrolysis reverting the conformational changes induced by it [151,152]. Upon taxol binding, GDP-bound MTs appear to have the same shape as GTP-bound MTs (they utilized the nonhydrolyzable GTP analog, GMPCPP, to reproduce a GTP-like state). However, as they stated, in order to clarify the molecular mechanism of this transition further structural analysis at near-atomic resolutions are required.



The same group, whose firstly obtained EC structure shed light into tubulin molecular details, solved the complex of Epothilone A and tubulin in the same way as they did before [151]. However, their results were in a contrast with the first tubulin-bound conformation of Epothilone A determined by solution state NMR experiments using a nonpolymerized soluble form of  $\alpha/\beta$ -tubulin [153]. For this reason, the binding mode of Epothilone A remained elusive until recently, when a high-resolution X-ray crystal structure of the drug in complex with a “curved”  $\alpha/\beta$ -tubulin was determined [135] showing clear differences from the previously proposed one (Figure 16). Another proof came from the results of Diaz and co-workers obtained by Saturation Transfer Difference (STD)-NMR analysis [154], which were in agreement with the X-ray model.

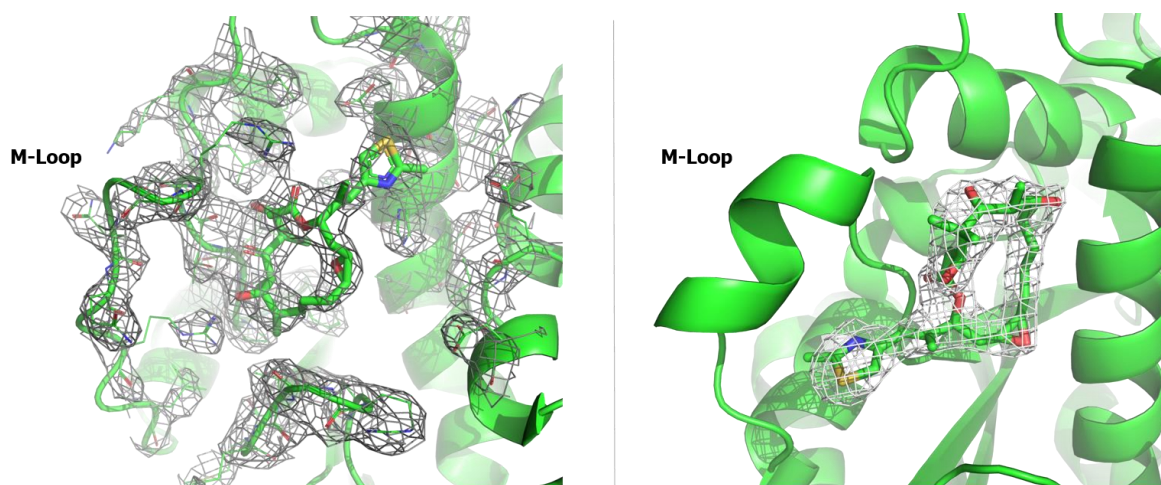


Figure 16 – Comparison between the originally proposed EP binding mode (on the left, PDB ID: 1TVK) and the recently one obtained through X-ray (on the right, PDB ID: 4I50). It is possible to note how the EC model lacks the necessary resolution to unambiguously assign the position of EP atoms, as is the case of the X-ray structure.

In light of these discordant results, a number of questions regarding taxol binding mode arise: does the zinc-induced tubulin sheets represent a good model for the straight conformation of tubulin? How accurate is the taxol binding mode proposed in 1998?

At the time of writing, there are still no X-ray structures available to confirm or confute the EC structure. New insights can only come from the recent advances in the cryo-electron microscopy field. It is now possible to reach near-atomic resolutions that can help in portraying the effects of MSAs on the overall MT structure (a recent work from Nogales and co-workers, through the use of cryo-EM reconstructions obtained at 3.9-4.2 Å, showed the

effects of taxol and peloruside on the MT lattice [155]). Furthermore, traditional docking techniques that rely on inaccurate crystallographic data, cannot deal with such a flexible binding site without incurring in the same mistakes [156].

For these reasons, exploring the possibilities offered by the recent advances in the dynamic docking field could be useful in shedding light on the mechanism of recognition of taxane-site ligands.

## **2.6. Aims of the Thesis**

Despite MDAs and MSAs have proven to be successful in the clinic treatment of a number of different type of cancers, it is not clear why some of them, are so effective against certain types of tumors, but basically useless against other types. For example, Taxol is routinely used against ovarian, mammary and lung tumors but has no effects on other solid tumors, such as kidney, or colon carcinomas. Clearly, some cells are more sensible than others to specific antimitotic drugs for a number of reasons that range from the expression of diverse protein isotypes to developed resistances [156].

However, their mechanisms of action is still argument of debate and there are also many doubts regarding their specific binding mode. These uncertainties derive from the intrinsic difficulties of dealing with a highly dynamic entity, as it is tubulin, a protein with the capability of undergoing a transition from a curved to a straight conformation that is a consequence of three different forms it can adopt (free soluble dimers, oligomers and polymerized in microtubules). Besides, due to the inherent difficulties of obtaining crystal structures of the straight conformation of tubulin, the available data is mainly related to the curved inhibited conformation of tubulin and, as shown by the case of Epothilone A binding mode, this can lead to discordant results.

From a computational standpoint, dealing with protein flexibility is not easy at all. However, in the last years, a number of different methodologies have been implemented to bring the study of the molecular interactions to a new level. Molecular dynamics and the related enhanced sampling methods are everyday gaining more importance in drug discovery and development programs.

With this in mind, the aim of this project is that of study the binding process of taxane-site ligands from a fully dynamic perspective. Above all, for unveiling the way these ligands exert their stabilizing action on the microtubule, it is necessary to acquire a better

understanding of the interactions that are at the basis of the molecular recognizing mechanism of these ligands by tubulin. In order to do so, the behavior of the binding pocket will be studied by means of classical molecular dynamic simulations. The resulting information will be used for the set-up of a fully dynamical protocol through the use of the approach recently implemented in the MD-Binding technique. The protocol will then be tested against three different ligands (epothilone A, discodermolide, and taxol) of the  $\beta$ -tubulin taxane-site to evaluate its feasibility and reliability. The relative performances, in terms of both strengths and weaknesses, against a traditional static approach, will be discussed.

## 3. Material and Methods

### 3.1. MD Simulations Setup

#### 3.1.1. System Preparation

The protein structural model for the simulations was taken from the experimental crystallographic structure accessible at the PDB code 4I4T. Simulations were performed on a single  $\alpha/\beta$ -tubulin heterodimer. Therefore, chains A and B, corresponding to the other heterodimer, E (stathmin), and F (TTL), were removed along with calcium and chloride ions. The GTP and GDP molecules and the  $Mg^{2+}$  ions found in the X-ray structure were included in the modelling. The same was done with the covalently bound zampanolide molecule. Water molecules involved in the interaction with the  $Mg^{2+}$  ions were maintained.

Residues of  $\beta$ -tubulin have been renumbered according to the UniProtKB [157] entry Q6B856. Therefore, the gaps in the numbering introduced by Lowe *et al.* [119] based on the alignment with the  $\alpha$ -tubulin are not present (see APPENDIX: Residue Numbering Correspondence for more details). The resulting structure possesses 440 out of 451 residues of the  $\alpha$ -tubulin (in respect to the UniProtKB P81947) and 431 out of 445 residues of the  $\beta$ -tubulin (UniProtKB Q6B856). Missing residues belong to the intrinsically disordered C-terminal domain and haven't been remodeled.

An initial guess for missing hydrogens and ambiguous protonation states have been evaluated through Schrödinger 2015-2 Protein Preparation Wizard [156] and chosen after a visual inspection. The histidine 227 of the helix H7 of taxane-site has been protonated at the epsilon position.

The protein structure has been solvated with TIP3P-model [156] water molecules in a truncated octahedron box using 12 Å as minimum distance between the protein and the box edges. The system has been neutralized by adding  $Na^+$  ions resulting in a total of 123776 atoms.

The force field Amber-ff99SB-ILDN [158] has been used for the protein. Parameters for  $Mg^{2+}$  ions and GTP/GDP molecules have been taken from the AMBER parameter database of Bryce Group [156].  $Mg^{2+}$  ions parameters have been developed by Allner *et al.* [159] whereas GTP/GDP parameters by Meagher *et al.* [160]. Accordingly, the O3 atom type has been manually inserted in the leaprc.ff99SBildn force field file.

The system has been assembled with the LEaP tool inside AmberTools 14 package [161]. Resulting coordinates and topologies files have been converted to GROMACS file formats with ACPYPE [162].

### **3.1.2. Plain MD Simulations**

Molecular dynamics (MD) simulations have been performed with GROMACS 4.6.7 [163]. The prepared system has been minimized with 5000 cycles of the steepest descent algorithm. At this point, the system has been equilibrated in four different stages using the V-rescale thermostat [164] and the Parrinello-Rahman barostat [165] with a relaxation time  $\tau$  of 2 ps and 0.1 ps, respectively, to keep the system at the desired target temperatures and pressure. The first three stages were run for 100 ps each, in the NVT ensemble with an integration time step of 1 fs, at increasing temperature values of 100 K, 200 K, 300 K, two of which with the backbone heavy atoms harmonically restrained with a force constant of 1000 kJ/mol/Å<sup>2</sup>. The last stage was run in the NPT ensemble for 1 ns with an integration time step of 1 fs. Bonds involving hydrogen atoms were restrained with the LINCS algorithm [166]. A short-range nonbonded cut-off of 9 Å was applied, whereas long-range electrostatics were treated with the particle mesh Ewald (PME) method [167]. Periodic boundary conditions (PBC) were applied.

After the equilibration stage, a 100 ns-long classical MD simulation has been conducted with an integration time step of 2 fs in the NPT ensemble at a target temperature and pressure of 300 K and 1 atm.

## **3.2. Binding Pocket Analysis**

Pockets were identified and tracked during the MD simulations by means of the Pocketron [168] module implemented in BiKi Life Sciences. For the analysis, the  $\alpha$ -tubulin subunit, the nucleotides, the ions, and the water molecules were removed. Small and large probe radii were set to 1.4 and 3 Å, respectively. Additionally, the minimum volume of a detectable pocket was set to the equivalent of 5 water molecules.

Static pockets and their relative “points of entrance” were evaluated using NanoShaper [169] using the same setting as for pocket tracking analysis.

### **3.3. MD-Binding Protocol Setup**

#### **3.3.1. Ligands Parametrization**

##### **3.3.1.1. Epothilone A**

Atomic coordinates of Epothilone A (EP) were taken from the PDB ID 4I50 corresponding to the X-ray structure of the T2R-TTL-EP complex. The electrostatic potential has been computed at the ab-initio HF/6-31G\* level theory with NWChem [168], as implemented in BiKi Life Sciences in its Residue Parametrization procedure, imposing a net charge of 0. The general amber force field (GAFF [170]) was employed to parameterize the ligand, while partial charges were fitted using the restrained electrostatic potential (RESP) procedure (via Antechamber of AmberTools 14), following the implementation of BiKi Life Sciences.

##### **3.3.1.2. Discodermolide**

No X-ray structure of Discodermolide (DDM) was available at the beginning of this study. The isomeric SMILES (simplified molecular-input line-entry system) of (+)-DDM, corresponding to the PubChem CID 643668, was used to produce an initial guess of the 3D structure by means of the LigPrep utility of Schrödinger Suite 2016-4. The structure has been then treated as reported in refs [171,172]. The lowest energy geometry has been obtained by conformational search with the MacroModel software (Schrödinger Suite 2016-4). The force fields AMBER\* and OPLS3 have been used. As searching and minimization algorithms, the mixed torsional/low-mode sampling method and the truncated Newton conjugate gradient procedure have been used, selecting 40000 structure generation steps (1000 per bond) and an energetic cutoff of 29.5 kJ/mol.

This obtained conformation has been further optimized at the ab-initio HF/6-31G\* level theory with the Gaussian03 code [173] and relative electrostatic potential has been evaluated. As for EP, the general amber force field (GAFF [167]) has been employed to parameterize the ligand DDM, while partial charges have been fitted using the RESP procedure via Antechamber of AmberTools 14.

##### **3.3.1.3. Taxol**

Taxol (TAX) structure has been prepared using the same procedure of DDM. The isomeric SMILE used corresponds to the PubChem CID 36314. The only difference is relative to the QM optimization. Before the optimization at the HF/6-31G\* level theory, the

initial lowest energy structure derived from the conformational search has been optimized at the semi-empirical method AM1 with the Gaussian03 code.

### **3.3.2. Ligand(s) Positioning**

The three ligands were placed nearby the binding site using the Reside Placement tool of BiKi Life Sciences. The binding site pocket has been visually individuated through VMD 1.9.3 [174] after the static pocket analysis of the MD frame whose pocket volume resulted the highest throughout the entire simulation course (for the relative analysis, see paragraph 3.2). The relative pocket entrance file has been used and, for the ligands, four different starting points and four random rotations were selected.

The last frame of the MD simulation was selected as protein model; concerning the ligands, respectively were used the X-ray model and the optimized conformations, for EP and for DDM and TAX.

### **3.3.3. Systems Preparation and Equilibration**

The resulting 16 systems for each ligand were prepared using the automated protocol implemented in BiKi Life Sciences. The first set of systems consisted each of a dimeric  $\alpha\beta$ -tubulin and one EP molecule. The subsequent set of systems included only the  $\beta$ -tubulin monomer and one of the ligands of interest among EP, DDM, and TAX.

System boxes, prepared through BiKi Life Sciences were then minimized and then equilibrated in five different stages, similarly to the tubulin dimeric system described in 3.1.2. The first two stages consisted of a total simulation time of 200 ps in the NVT ensemble at a temperature of 100 K and 200 K respectively. The equation of motion has been integrated every 1 fs and the heavy atoms of ligand and protein backbone were restrained. The other three stages, instead, were run with a time step of 2 fs. The third and the fourth, 100 ps of simulation each, in the NVT ensemble at 200 K and 300 K, with no restraints on protein backbone. The fifth stage was run in the NPT ensemble for 1 ns. In all the equilibration stages ligand heavy atom have been restrained.

### **3.3.4. MD-Binding Simulations**

MD-Binding Simulations were run with the GROMACS 4.6.1 MD engine patched with PLUMED2 [175], ad-hoc modified by BiKi Life Sciences. A series of different settings were used to validate the protocol and for productive simulations, which are summarized in Table

4. All the simulations were run in the NPT ensemble at 300 K and 1 atm, with an integration time step of 2 fs.

Reference (Setup)	Ligand	Simulation Time (ns)	K_GAIN	MAX_K	Switch-off residue(s) <sup>2</sup>	Attractive Residues <sup>2</sup>
<b><math>\alpha\beta</math>-tubulin heterodimer</b>						
1	EP	15	0.4	NA <sup>1</sup>	710, 712	463, 655, 667, 668, 670, 671, 710, 712, 713, 714, 719, 722, 723, 724, 798, 799, 801
2	EP	15	0.4	NA <sup>1</sup>	710, 712	463, 655, 667, 668, 670, 671, 710, 712, 713, 714, 798, 799, 801
3	EP	15	0.4	NA <sup>1</sup>	712	655, 664, 668, 710, 713, 714, 719
4	EP	15	0.4	NA <sup>1</sup>	712	655, 664, 667, 710, 712, 714, 719
5	EP	15	0.4	NA <sup>1</sup>	712	655, 667, 710, 712, 714, 719
6	EP	10	0.4	NA <sup>1</sup>	712	655, 667, 710, 712, 714, 719
7	EP	10	0.1	0.001	712	655, 667, 710, 712, 714, 719
8	EP	15	0.1	0.001	712	655, 667, 710, 712, 714, 719
9	EP	10	0.2	0.001	712	655, 667, 710, 712, 714, 719
10	EP	15	0.2	0.001	712	655, 667, 710, 712, 714, 719
11	EP	10	0.3	0.00001	712	655, 667, 710, 712, 714, 719
12	EP	15	0.3	0.00001	712	655, 667, 710, 712, 714, 719
<b><math>\beta</math>-tubulin monomer</b>						
13	EP	15	0.3	0.00001	272	215, 227, 270, 272, 274, 279
14	DDM	20	0.3	0.00001	272	215, 227, 270, 272, 274, 279
15	TAX	30	0.3	0.00001	272	215, 227, 270, 272, 274, 279

<sup>1</sup> NA stands for Not Available. The first version of the algorithm did not have this feature.

<sup>2</sup> Residue number as in the prepared system structures. They differs from the numbering adopted by Lowe *et al.* [119] for the reasons explained in the paragraph 3.1.1. See the appendix for Residue Numbering Correspondence.

Table 4 – Different setups used for MD-Binding simulations used in this work



### **3.3.5. Molecular Docking**

#### **3.3.5.1. Traditional Static Docking**

The software Glide [173] was used to test traditional molecular docking capabilities at reproducing the binding mode for the T2R-TTL-EP complex. The receptor grid was generated using, as protein conformation, the last frame of the classical MD simulation of the  $\alpha\beta$ -tubulin heterodimer. The grid center was defined as the centroid of residues His667 and Pro712. Default values were used for vdW radii and electrostatic scaling factors. Protein was treated as rigid body and ligand as flexible. For comparison purposes, both Glide scoring functions, GLIDE SP (Standard Precision) and GLIDE XP (Extra Precision), have been tested in different runs. It has been select to return 20 poses as output of docking calculations (the default is 1).

The ligand structure used has been previously prepared with the LigPrep utility of Schrödinger Suite 2016-2.

#### **3.3.5.2. Induced-Fit Docking**

Induced-Fit docking module implemented in Schrödinger Suite 2016-2 was tested against the same protein conformation as for the traditional static docking. The “enhanced” protocol has been selected. It consists of an initial Glide SP docking using a softened potential (vdW radii scaling factor of 0.5) followed by a Prime [176] Refinement (i.e. the minimization of the side-chains of receptor residues) of the obtained complexes and a final redocking with the default values for the scaling factors. The grid center was defined as for the traditional docking. For the minimization step were included all the residues at 6 Å from the ligand.

### **3.3.6. Scoring of MD-Binding Poses**

To evaluate MD-Binding poses at the end of the simulations water molecules were removed. As suggested by Glide user manual, when selecting the “minimize only” docking mode for rescoring purposes, the GLIDE SP scoring function has been chosen.

#### **3.3.6.1. Rescoring with MM/GBSA**

In order to rescore the poses obtained by the MD-Binding simulations, the protein and the ligand of last frames of each trajectory were extracted in two different files and subsequently fed to the Prime MM/GBSA module of Schrödinger Suite 2016-2. The

OPLS\_2005 force field was used along with the VGBSA implicit solvation model [176], the one implemented in Prime.

Protein flexibility has been evaluated minimizing all protein residues at 6.0 Å from ligand atoms.

### **3.3.7. Figure Preparation**

The figures of this work have been rendered either with the open-source version of Pymol 1.8.4.0 [177] or with VMD 1.9.3 [174]. Graphs were prepared using Gnuplot 4.6 [178].

## 4. Results and Discussion

### 4.1. Pocket Behavior Investigations

#### 4.1.1. Model Selection

The first step in every computational work is the choice of the starting structure. Due to the fact that this work is focused on the taxane-site, the chosen protein structure should not have any gap in this region. At the same time, in order to investigate dynamic docking capabilities in dealing with protein flexibility, it would be preferable to avoid any bias that could somehow help the implemented protocol.

The best-case scenario would be the possibility to have an apo structure of the protein of interest. Unfortunately, this does not apply to the tubulin case. As previously introduced, the taxane-site is composed by a number of stable helices (H6, H7, and part of H1), and two loops, one of which has a very peculiar shape that gives him the name, the so-called “M-loop”. It is highly flexible and no electron density associated to it is found in the crystal structure of the tubulin-stathmin-TTL-apo complex (PDB ID: 4IIJ). It can be stabilized either by the lateral interactions with other protofilaments within the MT or by the presence of certain MSAs bound to the taxane-site or to the peloruside-site.

These circumstances reduced the number of possible usable structures to the ones summarized in Table 5.

PDB ID	Resolution (Å)	Ligands	Reference
4I50	2.3	Epothilone A	[135]
4I4T	1.8	Zampanolide	[135]
4O4J	2.2	Peloruside A	[134]
4O4H	2.1	Laulimalide	[134]
4O4I	2.4	Laulimalide – Epothilone A	[134]
4O4L	2.2	Peloruside A – Epothilone A	[134]

Table 5 –  $\alpha\beta$ -tubulin X-ray structures where the M-loop is present.

Avoiding the complexes where epothilone A is present, for the reasons exposed above, the structure with the highest resolution is the one where zampanolide occupies the taxane-site. Despite its side-chain can be superimposed well to that of epothilone A, their macrolic cores are involved in a completely different set of interactions (the planes of the macrocycle are oriented at a  $\sim 90^\circ$  angle [135]). In addition, zampanolide is covalently bound to the NE2 (epsilon nitrogen) atom of His229 of the helix H7 (Figure 17). Indeed, the conformation of the amino acids of the binding site is certainly influenced by the bound molecule but, from the perspective of this work, this was going to be nothing more than another possibility to make things more fascinating.

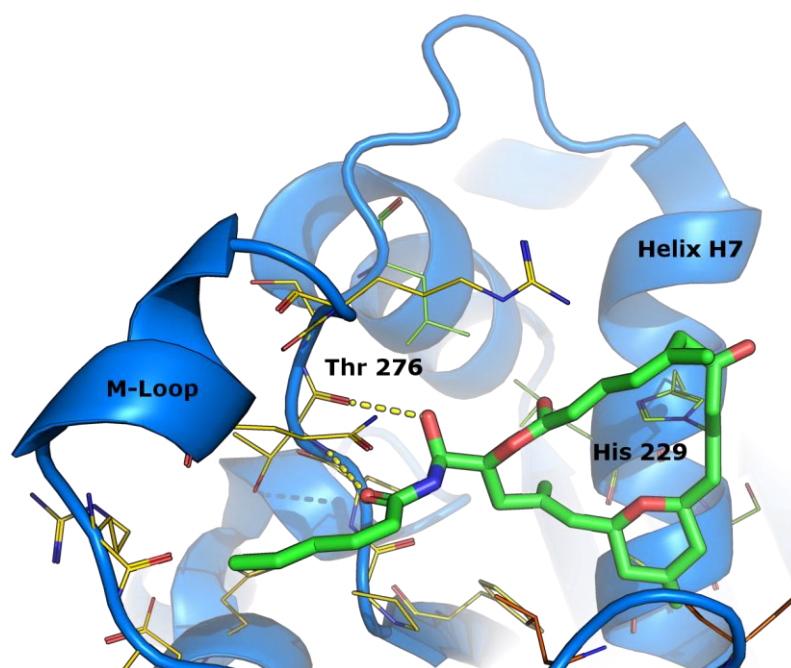


Figure 17 – Zampanolide binding mode as in its X-ray crystal structure (PDB ID: 4I4T – 1.8 Å). The ligand is covalently bound to the His 229 of Helix H7. Other than two hydrogens bond with the backbone of Thr 276, it is possible to identify many hydrophobic interactions.

We, therefore, decided to use the  $\alpha\beta$ -tubulin heterodimer of the T2R-TTL-Zampa complex, removing the tubulin tyrosine ligase, the other heterodimer lacking the ligand, and the stathmin-like domain of RB3. The same was done for the covalently bound zampanolide, imposing the His229 (i.e. the one involved in the binding with zampanolide) as epsilon protonated.

The resulting complex consisting of the two subunits and their relative nucleotide GTP or GDP, along with the  $Mg^{2+}$  ions and the water molecules coordinated with them, is highlighted in Figure 18.

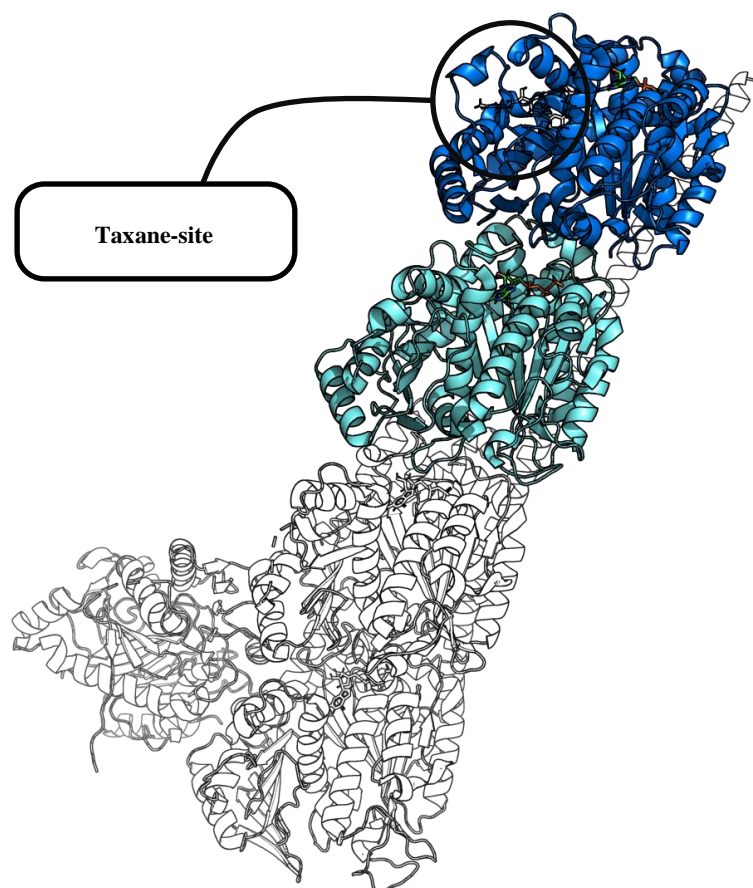


Figure 18 – The selected protein structure for this work. Chains removed from the PDB 4I4T are transparent, as it is the ligand zampanolide in the  $\beta$ -tubulin taxane-site.  $\beta$ -tubulin is colored in marine blue,  $\alpha$ -tubulin in light cyan.

There are no gaps in the selected structure, apart from a number of residues corresponding to the C-terminal tails of both subunits, which are known to be intrinsically disordered for the reasons explained in paragraph 2.3. These have an influence on MT dynamic instability but not on the internal stabilization of the monomers, thus have not been remodeled.

Nevertheless, with the aim of reducing the effects of the ligand in the taxane-site, we decided to run a 100 ns-long classical MD simulation of the complex from which select and retrieve a sort of apo-like conformation of the protein. In other words, this choice was mainly

driven by the preliminary condition of getting rid of any bias that could interfere with our simulations, both positively and negatively.

Before going into details, for a better understanding of the results, it is necessary to make an annotation regarding the residue numbering. Lowe *et al.* [118], aligning  $\alpha$  and  $\beta$  sequences, noticed that in order to match their secondary structures it was necessary to insert two gaps into  $\beta$  tubulin primary structure numbering. This is more evident at the level of the loop S9-S10, which is eight residues longer in the  $\alpha$ -tubulin subunit. As a matter of facts, this insertion occupies the taxane-site and explains why no ligands bind at the level of the  $\alpha$ -tubulin monomer. Anyway, for compatibility reasons with the MD engines that have been used in this work, we renumbered all the residues based on the correct amino acids sequence that is reported in the UniProtKB database (the corresponding ID is Q6B856). Therefore binding site residues numbering does not correspond to that reported in literature. For more details, it is possible to find a correspondence table in the APPENDIX.

#### 4.1.2. Classical MD-Simulation

In order to gain a deeper understanding of the structural and dynamic behavior of the taxol binding site, we ran one 100 ns-long MD simulation in explicit solvent, using the well-established atomistic force field Amber-ff99SB-ILDN [158]. From the initial conformation, the root mean square deviation (RMSD) of the C $\alpha$  carbons of the complex stabilizes after ~40 ns, and then oscillates around 1.8 Å for the remaining time of the simulation (Figure 19A). Considering the two subunits separately, the RMSD of the  $\alpha$ -tubulin stabilizes after ~40 ns, whereas the one of the  $\beta$ -tubulin stabilizes in a much shorter time (~10 ns). After this period, the two remain stable around values of ~ 1.5 and 1.25 respectively (Figure 19B-C).

It is worth noting that some residues of the intrinsically disordered C-terminal tails are still present in both subunits and their contributions must be carefully taken into account when computing the RMSD. In order to do so, we calculated the RMSD of the C $\alpha$  carbons of both C-terminal domains with and without considering these residues (Figure 20). As expected, the overall shape of the curves is very different, showing a remarkable impact of just few residues (five for the  $\alpha$ -tubulin and seven for the  $\beta$ -tubulin) on the result, which is consistent with the experimental data available [124]. We, therefore, decided to remove them from the subsequent analysis in order to provide a much clearer picture of protein stability.

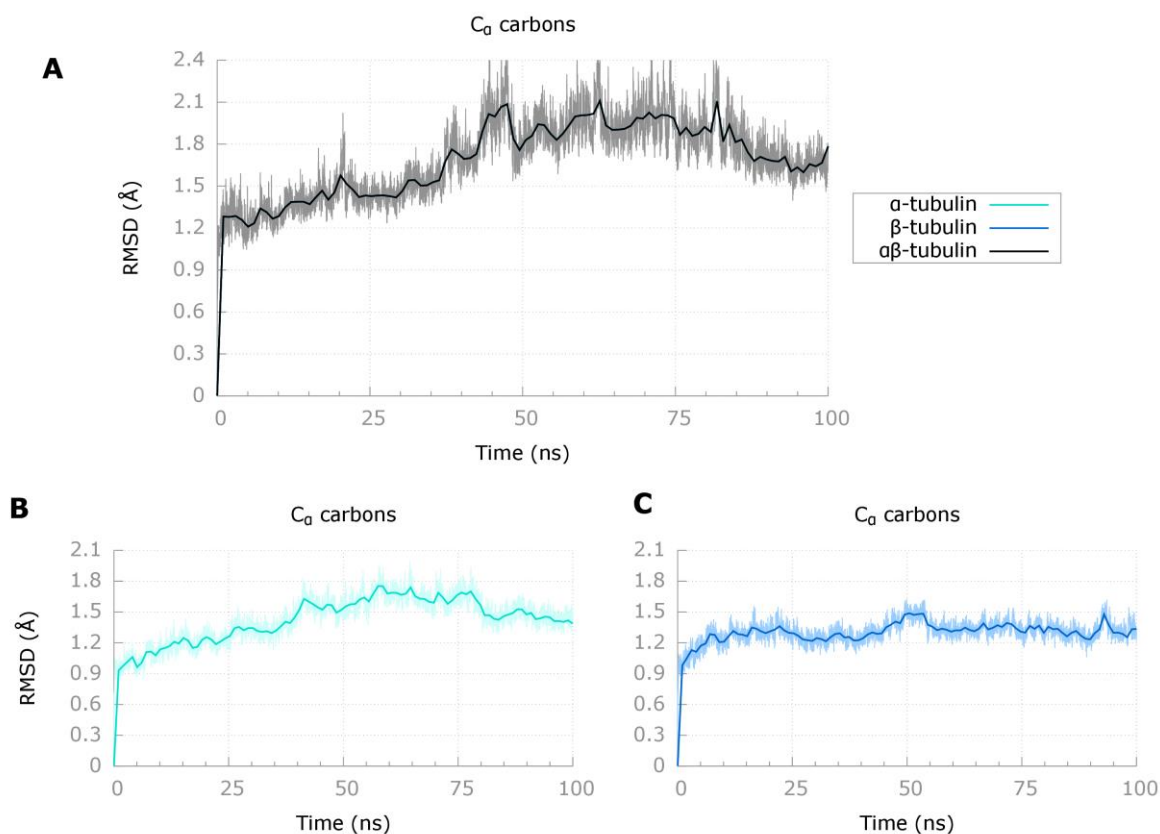


Figure 19 - Calculated RMSD of the  $\alpha\beta$ -tubulin heterodimer (A),  $\alpha$ -tubulin (B), and  $\beta$ -tubulin (C) subunits from their initial X-ray structure plotted as a function of time. The lightest lines represent the effective sampling of the RMSD during the entire simulation, whereas the darkest ones are smoothed in order to cut-off the noise.

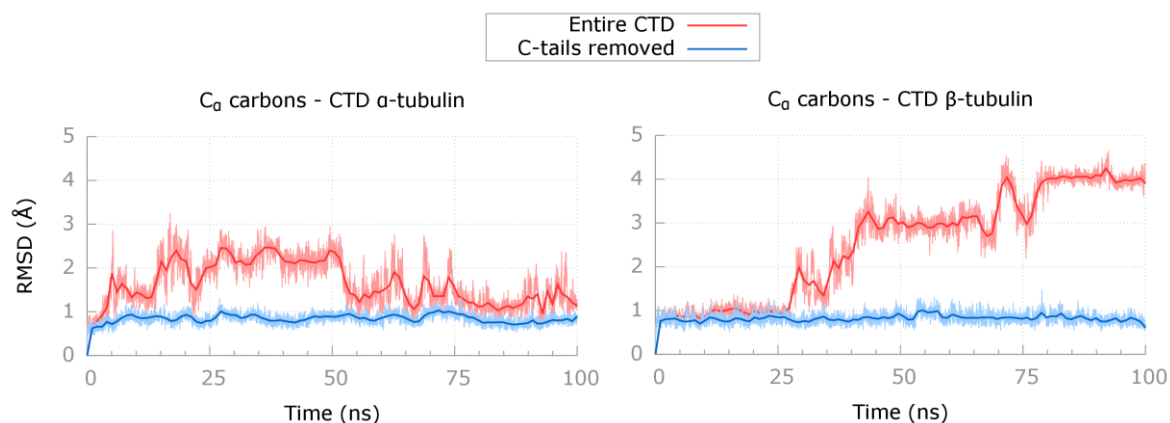


Figure 20 - Calculated RMSD of the C-terminal domains (CTDs) of both subunits from their initial X-ray structure plotted as a function of time. The red lines represent the RMSD computed taken into account the contribution of the entire CTD, while the blue ones have been obtained removing the last disordered residues.

Following this observation, we reevaluated independently the RMSD of the C $\alpha$  carbons of the three functional domains (see Table 2 and the Appendix for more information) of both subunits. The outcome of the analysis shows an even more reassuring result. Especially when considering the  $\beta$ -tubulin monomer (i.e. the one we are more interested into, seeing that it is where the taxane-site is located), the three domains are very stable along the entire MD simulation.

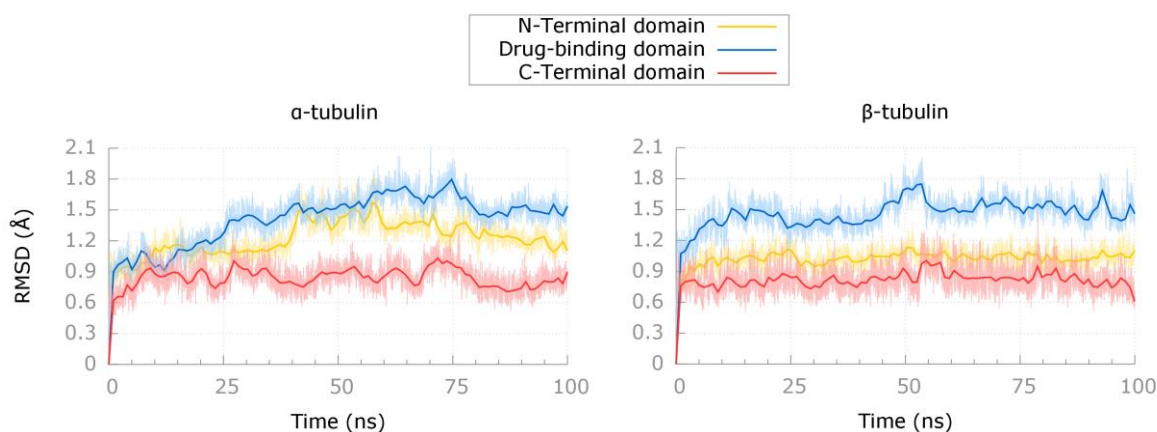


Figure 21 – Calculated RMSD of the C $\alpha$  carbons of the three domains from their initial X-ray structure plotted as a function of time. On the left, the RMSD associated to the  $\alpha$ -tubulin subunit. On the right, the one of the  $\beta$ -tubulin monomer.

These observations, taken together, indicate that the simulation was long enough to relax and equilibrate each tubulin subunit within the heterodimer, giving the ground for the next studies. In this perspective, we computed the root mean square fluctuations (RMSF) of  $\beta$ -tubulin C $\alpha$  carbons with the aim of identifying the most flexible regions of the protein. Again, the shows a situation that is in line with the experimental data. Only two regions show an RMSF of around 2.5 Å. One, corresponding to a small loop in the N-terminal domain (residues 54-58), and another one that lies in the drug-binding domain (residues 275-284) and relates to the M-loop that is part of the taxane-site. The latter, once loss the stabilizing effect of the interaction with the zampanolide side-chain, started to oscillate vigorously as expected (the X-ray structures of the apo protein do not have density associated to this loop).



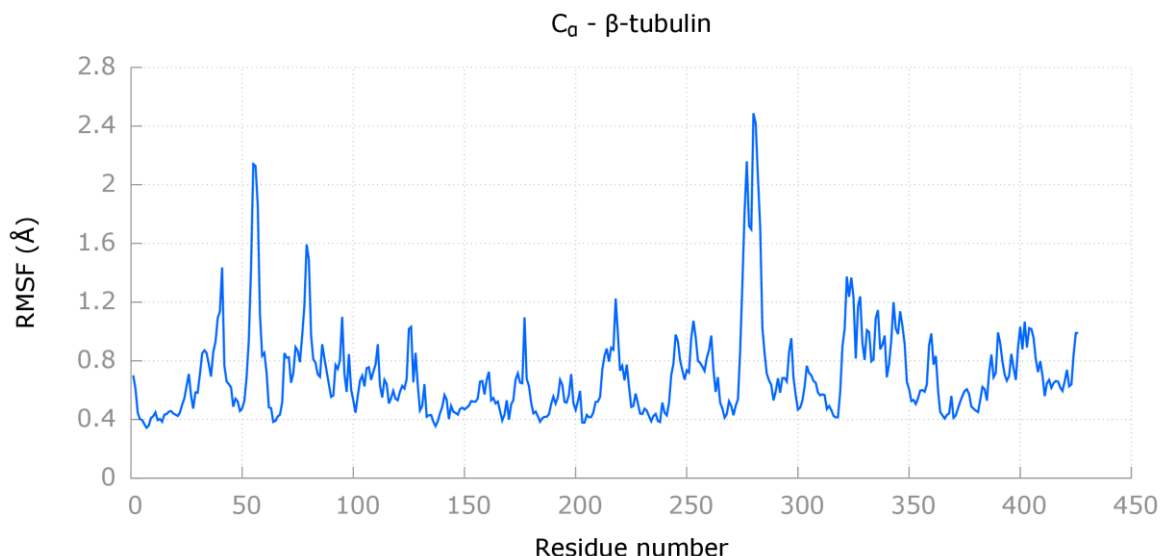


Figure 22 – Calculated RMSF of the C $\alpha$  carbons of the  $\beta$ -tubulin subunit during the simulation. For clarity reasons, the data coming from the most flexible C-terminal tails (residues 426-431) has been removed.

#### 4.1.3. Tracking the Binding Pocket

Following the assessment of the overall structural stability of the  $\alpha\beta$ -tubulin heterodimer, we focused our attention on the binding site pocket. The tool pocketron [168] of the BiKi Life Sciences suite has been used to analyze pocket dynamics. Its underlying algorithm for pocket detection is based on the solvent excluded surface (SES) concept [179] (i.e. the Connolly–Richards surface [180], obtained by rolling a spherical probe of a specific radius over the van der Waals surface of the biomolecular system). Following its assumptions, pockets are identified by calculating the volumetric difference between the regions enclosed by two different SESs, generated using two different probe radii, as implemented in NanoShaper [169].

In our analysis, we selected the default values of 1.4 and 3 Å for the probe radii, and, additionally, the equivalent of 5 water molecules as cutoff value for the minimum detectable volume. These settings have been chosen in order to filter the output and retrieve only those pockets that can be considered potential binding sites. Moreover, to reduce the computational cost of the analysis, we pre-extracted 1000 snapshots of only the  $\beta$ -tubulin subunit from our 100-ns long simulations, at defined time intervals.

Starting from the first frame of the simulation, pocketron monitors the evolution of the initially identified set of pockets, storing the inherent volumes and the residues that constitute them. At the same time, it is able to give information regarding pockets interactions, tracking the “merging” and “splitting” events that occur throughout the analyzed frames. Essentially, it counts all the times that a particular pocket splits into two different ones or, from the other point of view, when two different pockets join together in a single one. This analysis lets to examine the influence that a particular pocket can have on another one, even a very distant one.

The selected parameters led to the identification of 36 pockets on the  $\beta$ -tubulin subunit, four of which have average volumes greater than 100 Å<sup>3</sup> (Table 6). Another interesting aspect that has to be evaluated is the persistency of these pockets throughout the simulation, which indicates the fraction of the total simulation time in which a particular pocket is present. Nine of them were present for more of 40 % of the simulation time. As expected, being populated by the GDP, the corresponding binding site (pocket ID 5) was present for the entire simulation. Remarkably, we recorded the presence of two pockets in a region to whom binds the microtubule-stabilizing agent colchicine (pocket ID 7 and 8 with a persistency of about 69 and 90 %, respectively). In this respect, it is worth noting that during the simulation the analysis revealed also the presence of the peloruside A binding pocket, albeit for only the 15 % of the total time. As a matter of facts, this is another evidence of the tremendous power of MD simulations.

The taxane pocket, or at least the main part of it (pocket ID 4), was stable for the 95.8 % of the simulation with an average volume of 363.54 Å<sup>3</sup>. Close to it, in fact, another smaller pocket with a persistency of the 46.3 % is formed by some residues of the M-loop and comprises that region in which can accommodate epothilone A side-chain. Considering the high flexibility of the loop, as shown by the RMSF analysis (Figure 22), the lower persistency value comes not unexpected. Moreover, it is also possible that in some of the analyzed frames the two pockets are merged in one single entity, hence the 40 % persistency value represents only that fraction of the simulation time in which the two pockets were split. To verify this hypothesis we analyzed the merge and splitting events recorded during the analysis.

Pocket ID	Average Volume (Å <sup>3</sup> )	Persistency (%)	Notes
1	19.70	20.1	
2	59.48	46.3	Taxane-site
3	39.68	40.8	
4	363.54	95.8	Taxane-site
5	344.49	100.0	GTP/GDP site
6	109.34	81.0	
7	86.97	68.8	Colchicine-site
8	81.85	73.8	
9	198.10	89.9	Colchicine-site
10	16.95	16.7	
11	4.85	6.0	
12	15.62	15.3	Peloruside-site
13	25.49	29.9	
14	20.04	24.8	
15	0.52	0.7	
16	41.53	38.2	
17	7.24	8.7	
18	3.65	3.5	
19	45.80	44.7	
20	6.28	6.9	
21	0.74	0.9	
22	3.34	3.9	
23	9.35	9.8	
24	1.77	2.3	
25	0.37	0.5	
26	2.74	3.2	
27	0.59	0.8	
28	6.89	7.4	
29	21.66	22.1	
30	0.92	0.9	
31	3.25	3.2	
32	1.20	1.4	
33	4.42	5.2	
34	0.14	0.2	
35	0.14	0.2	
36	5.59	5.9	

Table 6 – Identified pockets and relative average volume and persistency.

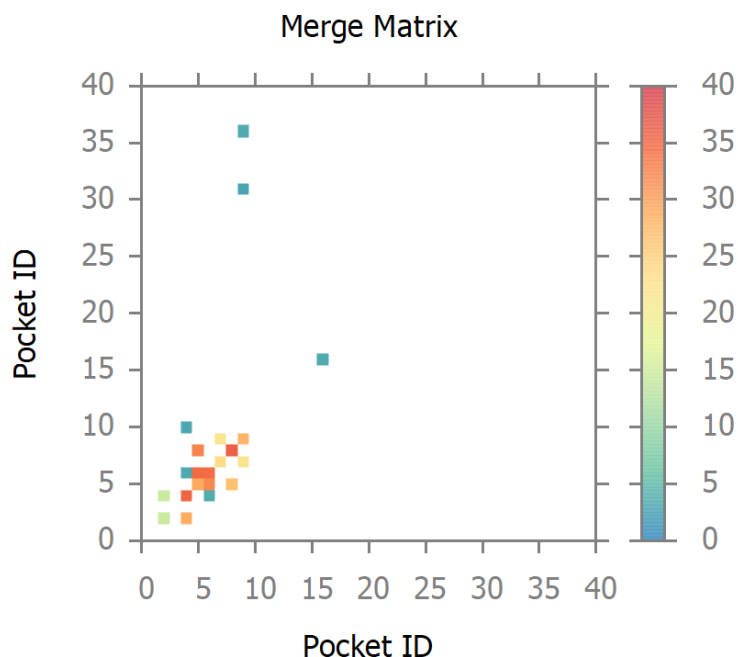
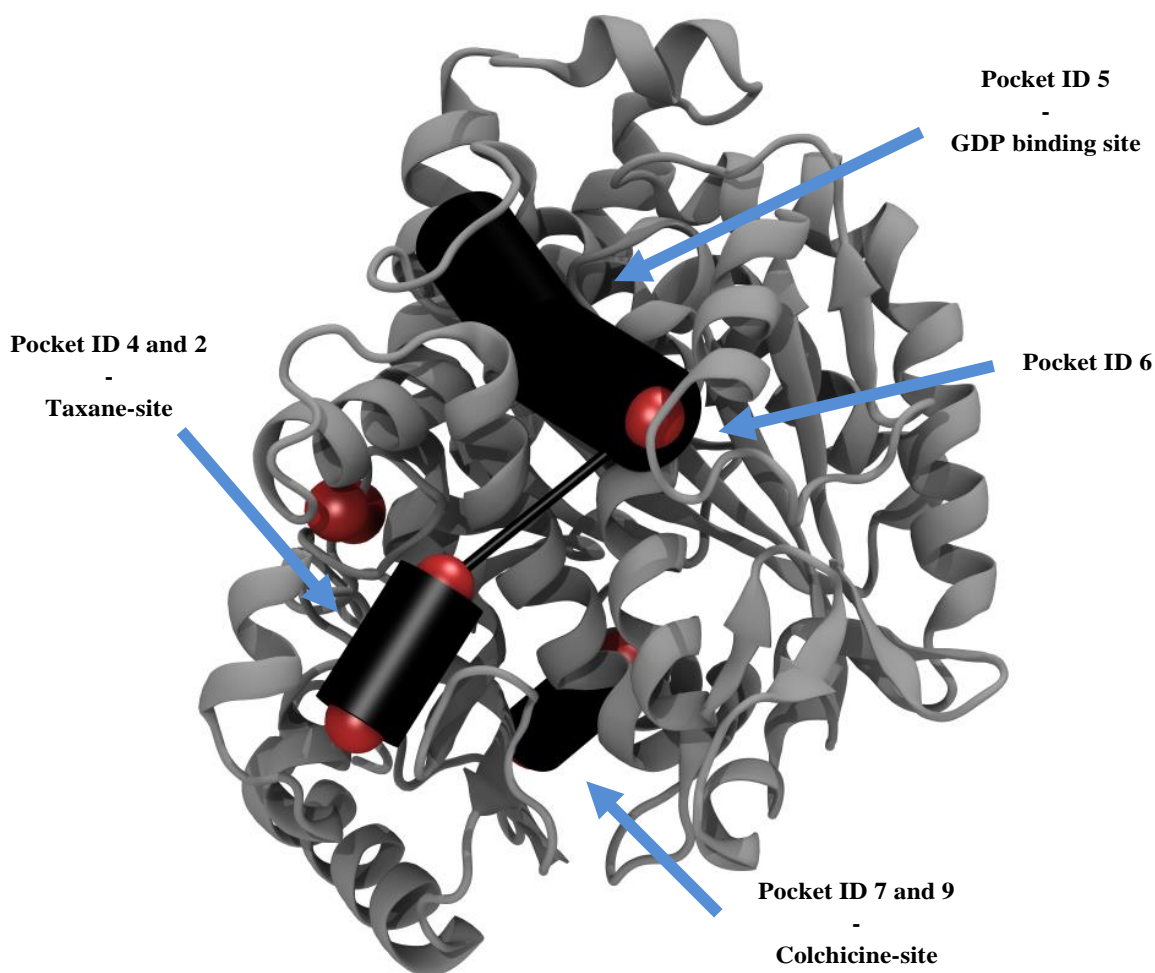


Figure 23 – Marge matrix of the 36 identified pockets. Values from 0 to 40 represent the merging probability between two pockets. For clarity reason, values over 40 have been removed.

We noticed that the two matrices (i.e. representing splitting and merging events) have similar values, hence we show only one of them. From the merge matrix represented in Figure 23 it is possible to observe a crosstalk between pockets ID 2 and 4, therefore our initial guess was somehow correct.

Nevertheless, the most interesting observation that comes out from this analysis is a series of merging events (albeit with a low probability) between pocket ID 4 and 6, the latter of which is strongly interconnected with pocket ID 5, that coincides with the GDP binding site. Once more, this is another evidence of the efficiency of our MD simulations. The experimental data, in fact, shows that when a GTP molecule is hydrolyzed to GDP, the induced conformational change causes instability within the MT, especially at the level of the lateral interactions between protofilaments [151], in which the M-loop seems to play a key role [155]. Moreover, MSAs like the drug taxol, are able to revert the conformational change induced by the hydrolysis of GDP [151] restoring and stabilizing these lateral interactions. In light of these considerations, it is reasonable to think that it should be communications between these pockets.

Pockets crosstalk can also be shown visually, as in Figure 24, where each node represents the geometric center of the atoms that form the corresponding pocket and the edges indicate how often two pockets exchange atoms.




---

Figure 24 - 3D network graph of pocket communication. The nodes represent pockets, and the edges indicate communication between two pockets. The thickness of each edge is directly proportional to the frequency of the merging events. Only pockets with a persistency greater than 40 % are shown.

We then focused our attention on the residues at the edge between the two pockets forming the taxane-site, pocket ID 4 and 2. These residues may act as a gate, mediating the opening and closing of the main pocket, thus modulating the entrance of the different ligands. Essentially, some atoms belonging to these residues can be involved in the formation of both pockets. Their relative persistency (for those that have it greater than 30%) have been

plotted in Figure 25, where it is possible to observe their relative contributions to the formation of both pockets. These are mainly residues belonging to the M-loop, such as Arg722, Gln716, Thr714 with long side chains or the smaller Ala723 and Pro712. Therefore, we had another confirmation that taxane-site is directly influenced by the flexibility of this loop.

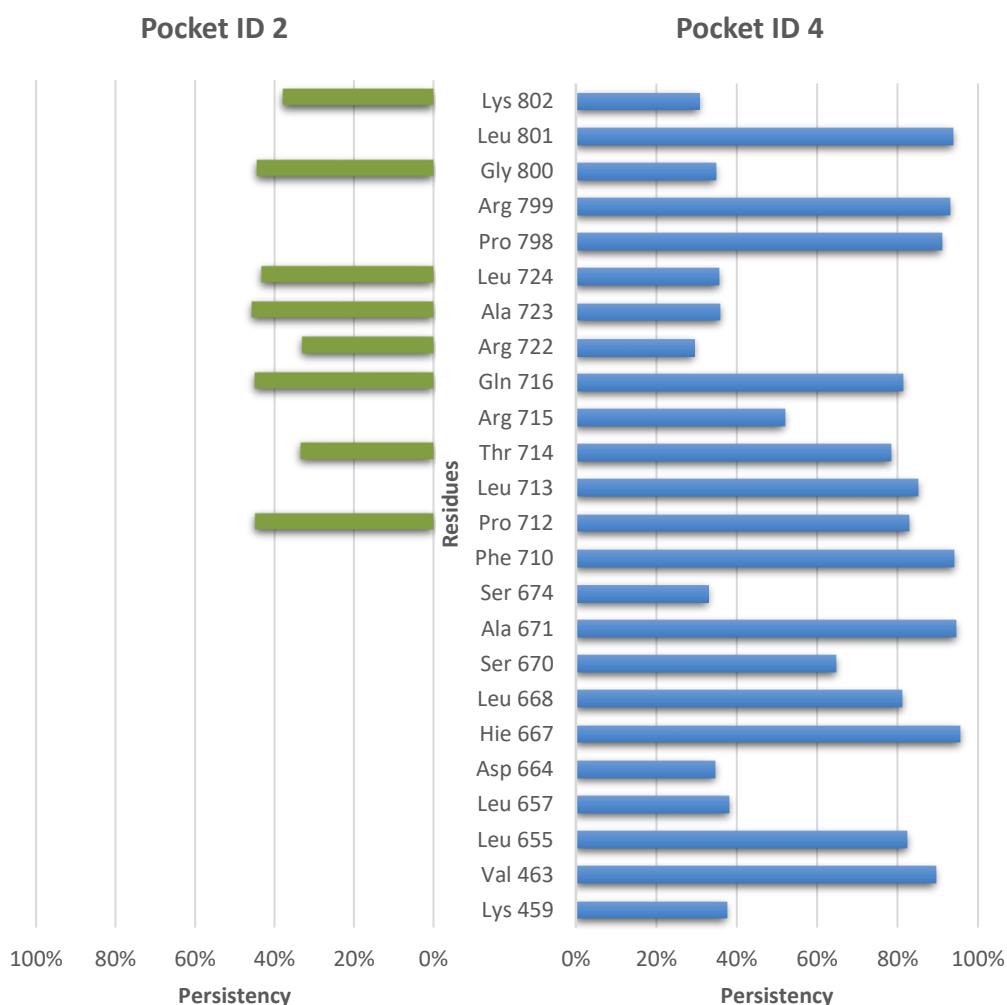


Figure 25 - Time persistency of residues that define the main taxane pocket (blue bars, pocket ID 4) and its correlated pocket (green bars, pocket ID 2). Only residues with a persistency greater than 30% are shown.

Following this consideration, we monitored the volumes of the two pockets along the simulations. We already knew that the main pocket (ID 4) is persistent for the major part of

the simulation (Table 6). However, the presence of pocket ID 2, and its volume thereof, is correlated to the merging events that occur throughout the simulation (Figure 26).

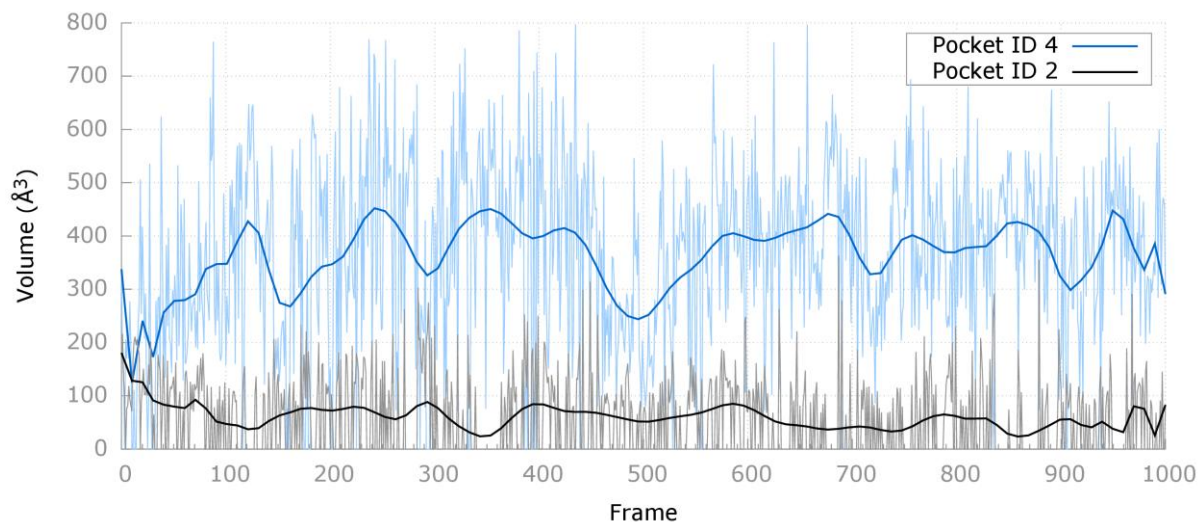


Figure 26 – Volume over time (expressed as frame number) of the two taxane-site pockets (ID 2 and 4). Light colored lines represent the effective sampling of the volume during the simulation, whereas the dark colored ones represent their approximation with a Bezier curve.

We noticed that when the volume of the main pocket increases, the other decreases, and *vice versa*. This behavior is more evident around the 500<sup>th</sup> frame when the main pocket almost vanishes or the frames 350<sup>th</sup> (35 ns of simulation) and 870<sup>th</sup> (87 ns) when it is the only one to be present. Specifically, the latter is related to a merging event recorded at the 863<sup>rd</sup> frame and the subsequent splitting at the 880<sup>th</sup>. This was the longest lasting period in which the two pockets were merged in one entity.

## 4.2. Considerations on the MD-Binding Protocol Setup

Going back to the traditional docking realm, the first aspect that has to be considered is the selection of the protein structure against which perform the docking calculation. Sometimes it may be right to select more than one protein model to work with, for example, if there are significant conformational shifts due to certain factors, such as the presence of other ligands, post-translational modifications, etc. [179].

For the reasons exposed above, conversely, a fully dynamical approach is free from this hurdle. Our 100-ns long simulations, other than for the evaluation of pocket behavior, was motivated by the need of getting rid of the binding site conformation induced by the ligand zampanolide. We wanted to put the MD-Binding protocol to the test with a challenging exercise and, because no complete X-ray structures were available, we regenerated a sort of “apo-like” protein status. Thus, we selected as starting structure for the subsequent enhanced sampling MD simulations, the last frame of our plain MD simulation.

Essentially, the MD-Binding protocol follows the workflow illustrated in Figure 27, and shares some of its steps to both docking and MD simulations.

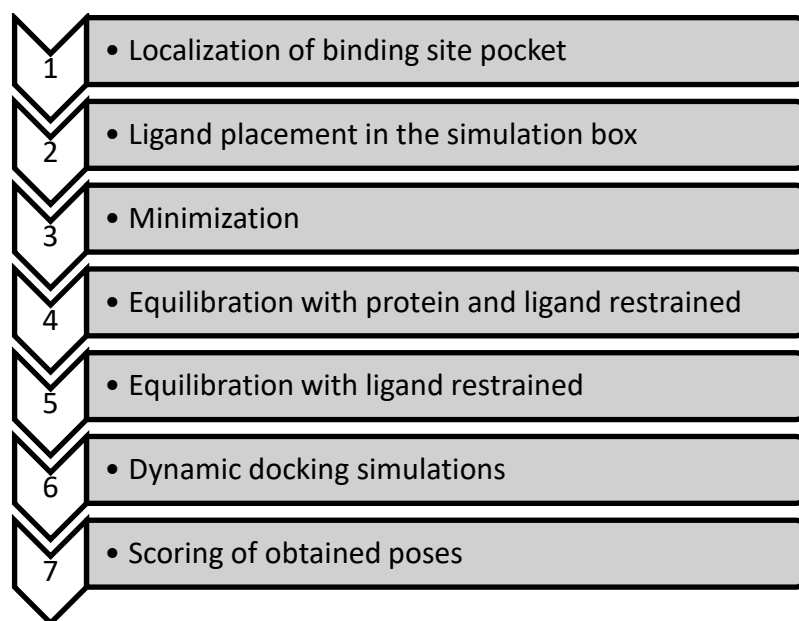


Figure 27 – MD-Binding workflow.



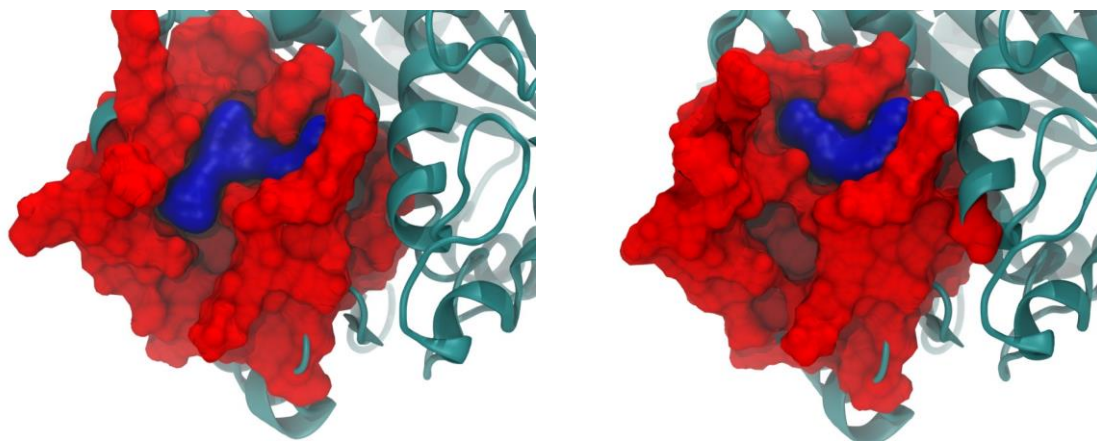
As a matter of facts, if from one end the MD-Binding technique overcomes the problems related to the selection of the conformation, from the other it introduces a series of delicate issues that are needed to be addressed (see paragraph 1.3). Firstly, the ligand molecule has to be placed in the simulation box in a way that it can reach its binding site in the user-selected simulation time (i.e. between 10 and 30 ns). Secondly, the method requires the selection and the tuning of a number of parameters:

- Attractive residue(s) - the residue(s) to whom the bias should be applied;
- Switching-off residue(s) - the residue(s) whose distance from the ligand determines the adaptiveness of the bias;
- Bias strength.

Regarding the former topic, we used the Residue Placement utility inside BiKi Life Sciences. The utility needs a “pocket entrance” file, generated by the NanoShaper static pocket detection tool. NanoShaper, in fact, identifies all the possible entrances to the given pocket and positions a set of points on each of them. In a second step, these points are clustered, and the centroid relative to each cluster is considered as a putative access door of the ligand to the pocket. Seeing that it depends from the shape of the relative pocket and that the taxane site is highly mutable through the entire simulation, we had to select which frame to employ for the static pocket analysis. According to the volume-over-time plot (Figure 26), we decided to select one of the frames belonging to the longest lasting period in which the two related pockets of the taxane-site were merged in one single entity. Specifically, we selected the 875<sup>th</sup> frame, the one in which we recorded the highest volume for the binding site pocket.

For comparison, Figure 28 shows the different shapes assumed by the pocket in the 875<sup>th</sup> frame (the one used for generating the pocket entrance file) and the last frame of our simulation (the one selected as starting point for the subsequent enhanced sampling MD simulations).

Once these access doors have been identified, the utility place a ligand molecule 10 Å away from each of them along the normal to the relative point, with a series of different random orientations. In our case, we identified 4 points of entrance and selected to use 4 random ligand orientations.




---

Figure 28 - Pocket volume (showed in blue surface) at different times during the simulation generated by NanoShaper. On the left, the 875<sup>th</sup> frame used for generating the pocket entrance file, and, on the right, the last frame of the simulation, where the taxane-site pocket is split into two different pockets, therefore pocket ID 4 volume is much smaller.

We then prepared the resulting structures according to the paragraph 3.3.3, restraining the ligand heavy atoms during the equilibration steps. We did this in order to avoid having it floating around the binding site, invalidating the simulations.

Regarding the rest of the parameters for the production runs, we based our settings on a step-by-step approach. Our first attempts were defined without using any *a priori* knowledge (Reference setups 1, 2 of Table 4). We used as attractive residues, the ones composing the taxane pocket at the end of the plain MD trajectory and, as switching off residues, Pro710 and Phe712. We imposed a simulation time of 15 ns and a K\_GAIN of 0.4, the defaults values suggested by BiKi Life Sciences. In our first set of simulations, we used the entire  $\alpha\beta$ -tubulin heterodimer.

We then tested a different sets of atoms for both attractive and switching-off sets of residues (Reference setups 3-5 of Table 4) and lowered the K\_GAIN parameter to a value as minimum as possible to reduce the bias strength and, at the same time, the influence of the bias itself on the simulations (Reference setups 6-12 of Table 4). Latest versions of the protocol implemented also a MAX\_K parameter in order to avoid a too abrupt introduction of biasing force at the beginning of the simulations. We tested two different settings for the MAX\_K parameter (Reference setups 10-12 of Table 4).

## 4.3. Protocol Validation

In line with any traditional docking workflow, the very first aspect to consider is if the implemented protocol is able to recover the experimental binding mode. We had to discriminate between the two different epothilone A binding modes, knowing that one of them is likely to be incorrect (the EC structure obtained at 2.9 Å, PDB ID: 1TVK) or remarkably different from the one adopted in free-tubulin heterodimers, the structure we are studying. In the best-case scenario, our protocol should be able to recover only the X-ray crystallographic pose (PDB ID: 4I50 – 2.3 Å) or, at least, to recover the X-ray pose a number of times higher than the EC one.

### 4.3.1. Traditional Docking Procedures

In this respect, before putting the MD-Binding protocol in place, we verified the capabilities of one of the major docking software currently available, Glide [181], to reproduce the X-ray pose using as protein conformation the one we previously selected. We tested both its scoring function, the standard precision (GLIDE SP) and the extra-precision (GLIDE XP). In the same perspective, knowing the importance of the taxane-site flexibility, we also tested the induced-fit module [23] implemented in the same suite (Schrödinger Suite 2016-2). For the induced-fit docking, we did not specify any particular residue but we selected to sample side-chains conformations of all the residues within 6 Å of any of the ligand atoms, using the “extended sampling” procedure.

The results did not come unexpected. None of them was able to recover both the X-ray pose and the EC one. Considering the conformation of the receptor we used (Figure 28, on the right), just from a purely visual inspection it seems clear that a molecule like the epothilone A cannot fit in such a narrow space. Even the specifically developed induced-fit module cannot deal with the large loop rearrangements of the binding site. Thinking on the way it works, the first step is always a traditional docking procedure, therefore, the ligand is completely flexible whereas the protein is kept rigid. Besides, despite the use of a considerable scaling factor for the vdW radii (0.5), it is unreasonable to think that the subsequent side-chains rearrangements could produce enough space for the ligand accommodation.

From another standpoint, it is worth noting that, using the protein conformation of the X-ray structure (4I50), the Glide algorithm succeeded in the self-docking test, it correctly

reproduced the crystallographic binding mode. Once again, this put the emphasis on the impact that the selection process of the protein conformations would have in traditional docking procedures, especially for those cases where large loop rearrangements are reported upon binding.

#### 4.3.2. The Two Epothilone A Binding Modes at Test

Proceeding with our studies, we wanted to assess the stability of the two crystallographic binding modes. In order to do so, we ran two 100-ns long plain MD simulations starting from the two crystal complexes, corresponding to the PDB ID 1TVK and 4I50. From the initial conformation, the RMSD of the ligand heavy atoms showed an evident different behavior for the two binding modes (Figure 29).

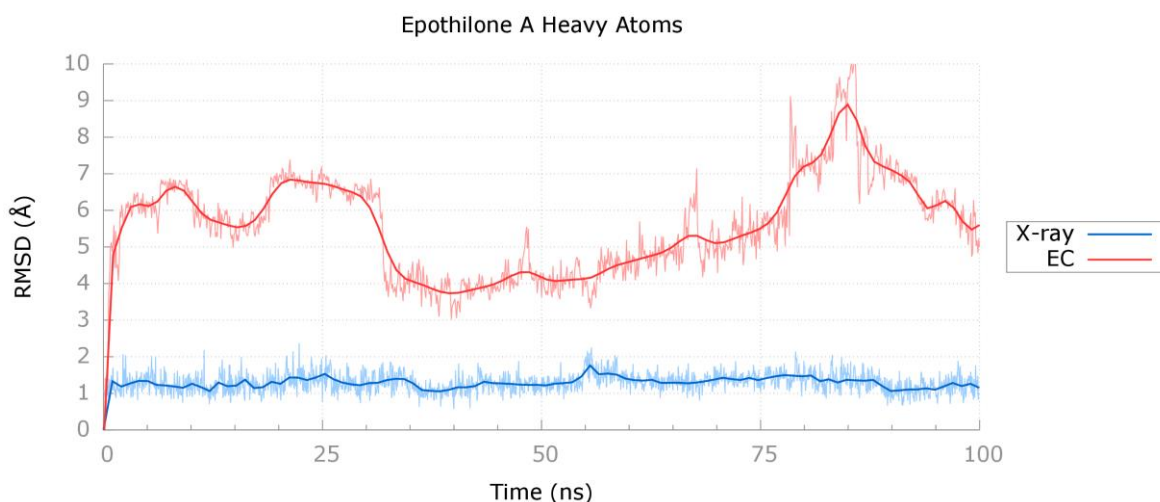


Figure 29 - Calculated RMSD of the ligand heavy atoms from their initial crystallographic structure plotted as a function of time, after fitting on binding site C $\alpha$  carbons.

Whereas the X-ray binding mode is stable for the entire simulation (the average RMSD value is 1.23 Å), the trend in the RMSD plot for the simulation that started from the EC crystallographic structure is typical of an unstable pose (the average RMSD value is 5.78 Å, with peaks of 10.46 Å). During the simulation, the macrolidic core of the ligand rotates around its center-of-mass (COM) and the side chain, being solvent exposed, does not make and durable interaction with the protein.

In the same context, it is interesting to note that the RMSF of the  $\beta$ -tubulin C $\alpha$  carbons in the simulation that started from the X-ray complex (Figure 30) is similar to that computed for the previously simulated apo-structure with a significant exception. The M-loop region (corresponding to residues 272-285) is now less flexible due to the stabilization effect of epothilone A (1.5 Å vs 2.5 Å. For comparison, see Figure 22 and Figure 30). Conversely, in the simulation that started from the EC complex, the M-loop is way less stable than both the two other simulations (5.5 Å vs 1.5 Å - Figure 30). Given the importance of the M-loop in the lateral interactions with other protofilaments, and the stabilization effect shown by epothilone A on the MT lattice, this can be considered another proof against the validity of the binding mode proposed by the EC density map.

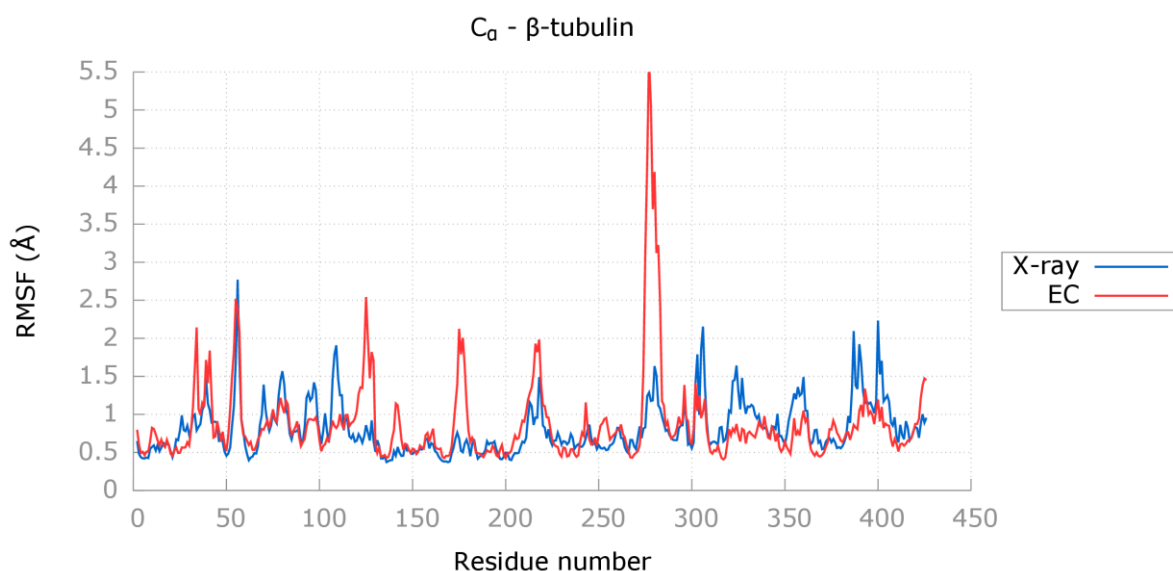


Figure 30 - Calculated RMSF of the C $\alpha$  carbons of the  $\beta$ -tubulin subunit during the simulations of the two complexes. For clarity reasons, the data coming from the most flexible C-terminal tails (residues 426-431) has been removed.

### 4.3.3. Dynamic Docking of Epothilone A

#### 4.3.3.1. Posing

We then moved to the analysis of the MD-Binding approach to dynamic docking. We ran our simulations through the BiKi Life Sciences interface and monitored the ligand's heavy atoms RMSD against the two crystallographic structures 1TVK and 4I50.

As stated in paragraph 4.2, we started with a generic approach, selecting all the residues of the binding pocket as “attractive residues” and the deepest among them, as “switching-off residues” (Figure 31 - Reference setups 1 of Table 4.). This kind of approach does not exploit any *a priori* knowledge of the system of interest, apart from the location of the binding site itself. From this perspective, looking at the single tubulin heterodimer in solution, the taxane-site looks more like a superficial binding site rather than an inner cavity, as it is the colchicine-site (Figure 13).

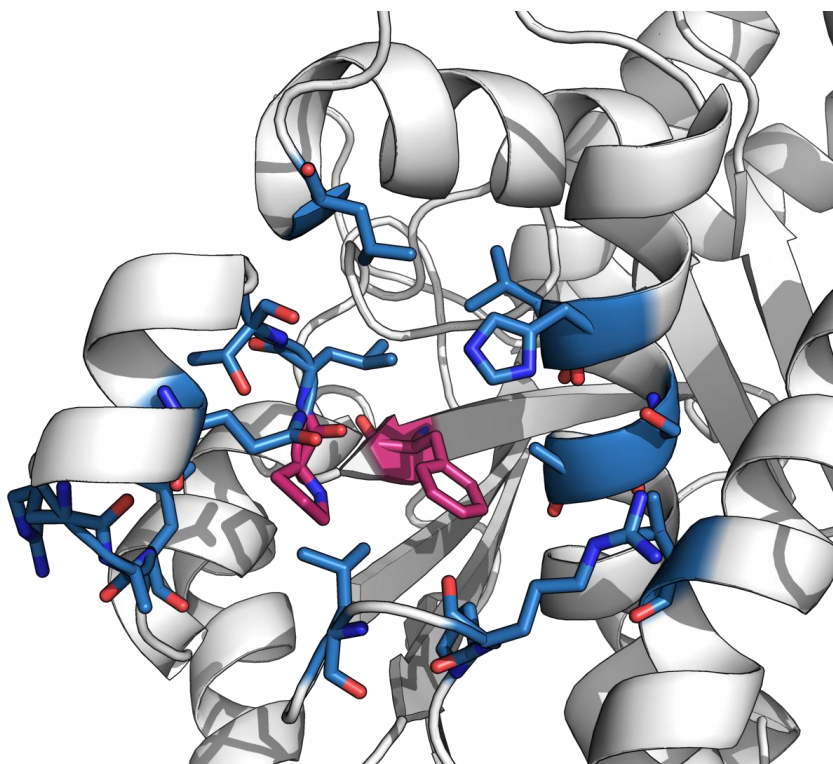


Figure 31 – Reference setup 1 of Table 4. Attractive residues are shown in blue, switching-off residues in magenta.

Our first set of simulations did not recover any of the crystallographic poses (Table 7) and we thought that it was due to our residues selection. Despite the adaptiveness of the methodology, it seemed to us that the amount of residues we selected attracted the ligand to various directions, producing a diffused effect of the virtual charges.

Run ID	Final RMSD against X-Ray Structure (Å)	Best RMSD against X-Ray Structure (Å)	Best RMSD against EC Structure (Å)
1	6.96	5.27	8.04
2	10.53	6.78	12.76
3	4.98	4.06	7.80
4	8.14	7.44	4.92
5	7.91	6.77	5.21
6	9.49	8.26	4.10
7	8.26	7.82	4.81
8	6.59	6.22	4.63
9	9.18	6.33	7.27
10	9.98	7.97	9.32
11	9.43	8.13	6.35
12	6.32	5.56	9.16
13	9.72	7.94	6.23
14	7.17	6.21	7.08
15	15.78	14.7	15.18
16	8.80	7.71	5.20

Table 7 – Computed RMSD of the ligand’s heavy atoms during the MD-Binding simulations run according to the parameters specified in Table 4 - reference setup 1.

Moreover, it is useful to remember that as soon as any atom of the ligand reaches any of the atoms of the switching-off residues (i.e. the pairwise distance between them should be less than 4 Å), the bias is gradually lowered and the simulation reaches a plain-MD like status. With this in mind, we decided to put a different rationale on the choice of the residues trying different setups. We looked at the pocket and, at the end, we imposed as switching-off residue only the Pro712, avoiding the unnecessary addition of the Phe710. The latter is positioned in a way that does not smooth the interaction with the ligand, therefore its switching-off control is not exploited. Additionally, we reduced the number of the attractive residues to six: Leu655, His667, Phe710, Pro712, Thr714, and Gln719 (Figure 32).

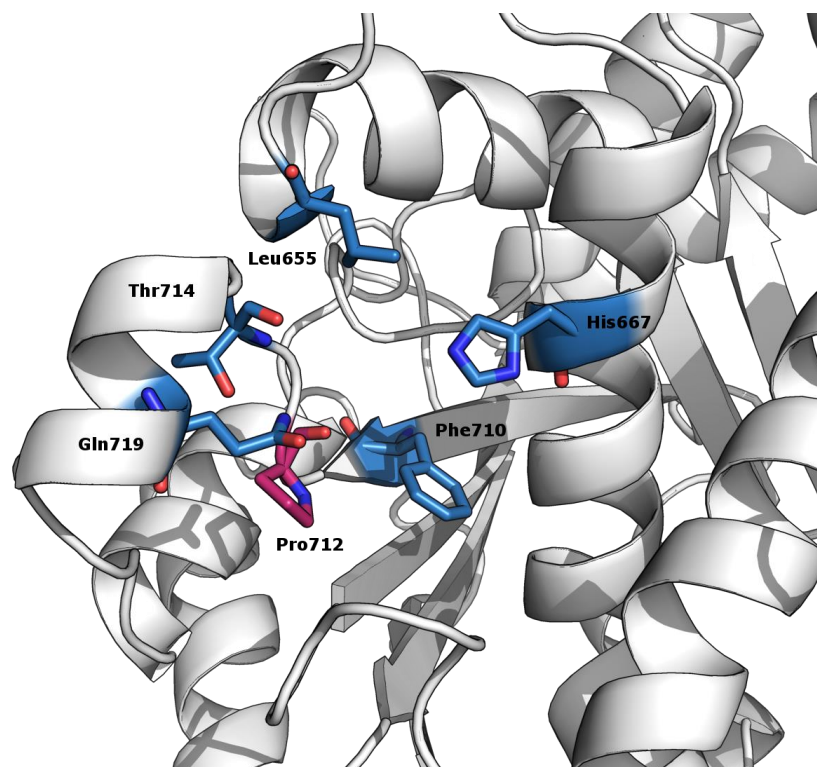


Figure 32 - Reference setups 5-12 of Table 4. Attractive residues (Leu655, His667, Phe710, Pro712, Thr714, and Gln719) are shown in blue, the switching-off residue Pro712 in magenta.

According to the crystallographic binding mode of PDB ID 4I50, some of these are the residues that make contact with epothilone A. We avoided the use of Asp664, despite its known interaction with one of two macrocycle hydroxyl groups because, being located at the surface of the protein, it has not been identified as part of the pocket by the NanoShaper software.

With these new parameters, in one out of the 16 different runs, we recorded an RMSD value as low as 0.69 Å for ligand's heavy atoms compared to the X-ray crystal structure (Table 8). We initially thought to have recovered the X-ray binding mode but, looking more carefully at the structure, we noticed some discrepancies. Despite the hydrophobic interactions with the pocket wall and the macrolidic part of epothilone A are maintained (Phe710, Pro712, Leu668, and Leu655), the M-loop helix induced by the MSA zampanolide is now less ordered, and the relative residues are in a different conformation. As illustrated in Figure 33, the h-bond between the side chain of Gln719 and the macrocycle hydroxyl group does not take place. Instead, Gln719 is involved in another h-bond with the backbone



of Leu655. The same can be said for the h-bond between the hydroxyl group of the Thr714 and the thiazole nitrogen of epothilone A lateral chain.

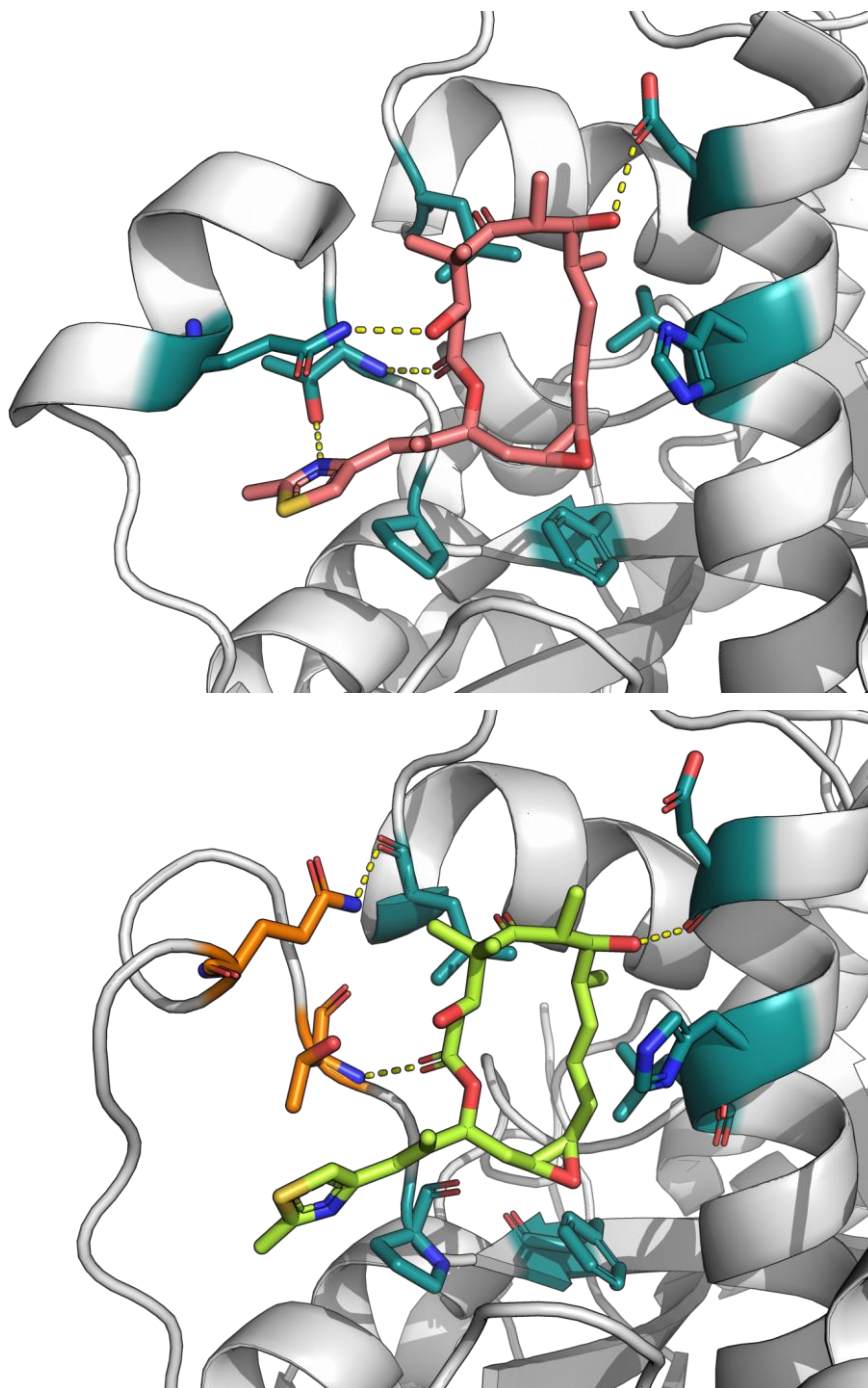


Figure 33 – Comparison between the crystallographic binding mode of epothilone A (top) and the one obtained by means of MD-Binding (bottom). Residues with a different conformation (Thr714 and Gln719) are highlighted in orange.

Run ID	Final RMSD against X-Ray Structure (Å)	Best RMSD against X-Ray Structure (Å)	Best RMSD against EC Structure (Å)
1	2.78	1.90	7.41
2	9.18	8.33	4.65
3	7.97	7.72	5.85
4	7.31	7.13	5.98
5	8.34	7.91	9.57
6	8.54	8.04	9.03
7	8.93	8.13	4.35
8	9.00	8.52	4.02
9	1.08	0.69	8.86
10	6.64	5.42	6.77
11	6.63	4.88	7.58
12	8.17	7.27	5.57
13	8.69	6.89	5.31
14	6.55	6.20	7.52
15	7.97	6.24	7.93
16	8.80	8.42	4.97

Table 8 – Computed RMSD of the ligand’s heavy atoms during the MD-Binding simulations run according to the parameters specified in Table 4 - reference setup 6.

In light of these results, we considered the possibility to reduce the strength of the imposed artificial bias. We, therefore, tested different values for the K\_GAIN (0.1, 0.2, and 0.3) parameter. At the same time, BiKi Life Sciences implemented a new adaptive parameter, MAX\_K, with the aim of attenuating the effect of the bias at the beginning of the simulations and avoiding to have a harsh acceleration for the ligand when it is in the bulk solvent.

With the K\_GAIN set to 0.1 and 0.2, we did not recover the X-ray structure. Conversely, we found the perfect balance with a K\_GAIN value of 0.3 and a MAX\_K value of 0.00001. With these settings, corresponding to the reference setup 11 and 12 of Table 4 that differs only for the simulation time (10 and 15 ns, respectively), succeeded in recovering the crystallographic binding mode. Besides, aware of the previous mistake, instead of monitoring only the RMSD of ligand’s heavy atom, we included in the computation also the backbone of the binding-site residues. This addition produced a better picture of the binding event (Table 9) letting us to properly identify the crystallographic binding mode (Figure 34).

Run ID	Final RMSD against X-Ray Structure (Å)	Best RMSD against X-Ray Structure (Å)	Best RMSD against EC Structure (Å)
1	7.44	7.15	6.01
2	9.20	7.35	11.0
3	9.85	9.50	9.32
4	7.42	2.77	8.08
5	1.61	1.13	7.99
6	7.07	5.87	7.73
7	8.58	6.00	5.39
8	7.03	4.38	8.17
9	8.59	6.78	5.71
10	8.79	7.35	5.47
11	5.02	3.86	8.15
12	7.43	6.27	6.85
13	8.90	8.36	6.93
14	9.83	9.28	8.78
15	8.97	7.29	3.41
16	11.89	10.46	11.73

Table 9 - Computed RMSD of the ligand's heavy atoms and the backbone of binding-site residues during the MD-Binding simulations run according to the parameters specified in Table 4 - reference setup 12.

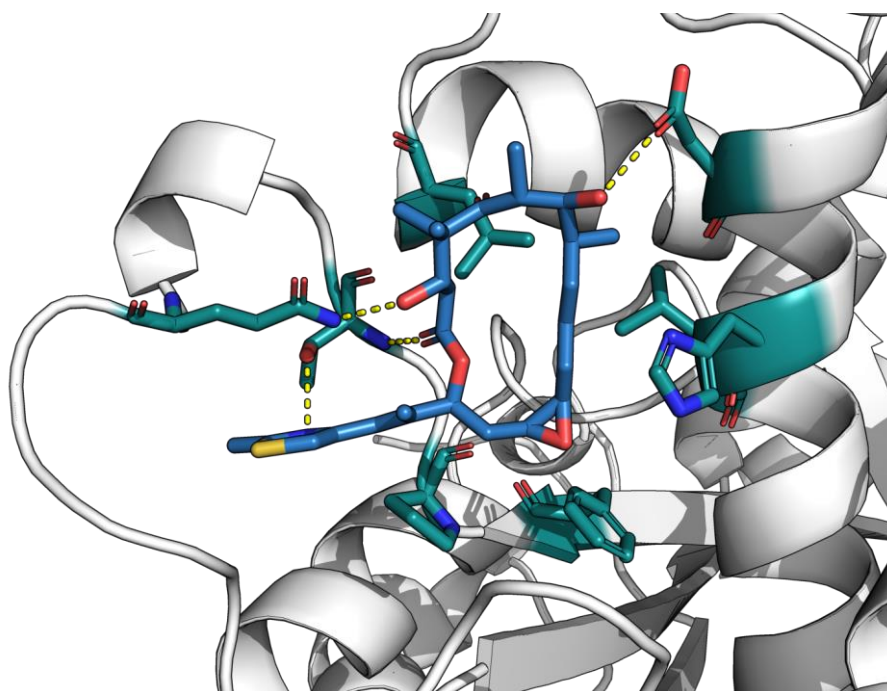


Figure 34 – The correctly reproduced binding mode of epothilone A with MD-Binding. Reference setup 16 of Table 4, Run ID 5. For comparison, see Figure 33 (top).

For comparison, Figure 35 shows the different evolution of the bias (dictated by the parameter  $k$ ) with the above-mentioned two different setups. The introduction of the MAX\_K value smooths the switching-on of the bias in the first part of the simulation and. Its effect is noticeable from the trend of the distance between the ligand and the switching-off residue. The ligand, in fact, moves slower toward its binding site following a more natural behavior.

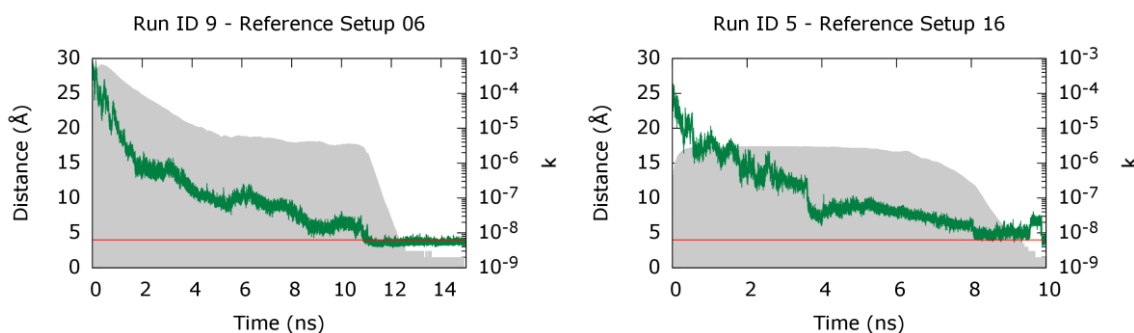


Figure 35 – Evolution of the bias (grey area) for two different MD-Binding setups. The green line represents the distance between the ligand and the switching-off residue Pro712 monitored along the MD simulations. The red line is the imposed threshold value of 4 Å that controls the bias switch-off.

#### 4.3.3.2. The importance of the His227

An interesting observation that came out looking at the simulations is the role played by the His227. We noticed that in many different runs, when the ligands struggled to reach the crystallographic binding mode, the histidine side-chain was located in a way that prevented the proper positioning of the epothilone A macrolidic core. Accordingly, despite the interaction with the backbone and the side chain of Thr274 are maintained, allowing the correct positioning of the epothilone A lateral chain, the macrocycle assumes a different conformation that is unsuitable for establishing the hydrogen bonds with the side chains of Asp224 and Gln279. The latter is central to the stabilization of the small M-loop alpha helix, which, for this reason, is almost completely unfolded (Figure 36).

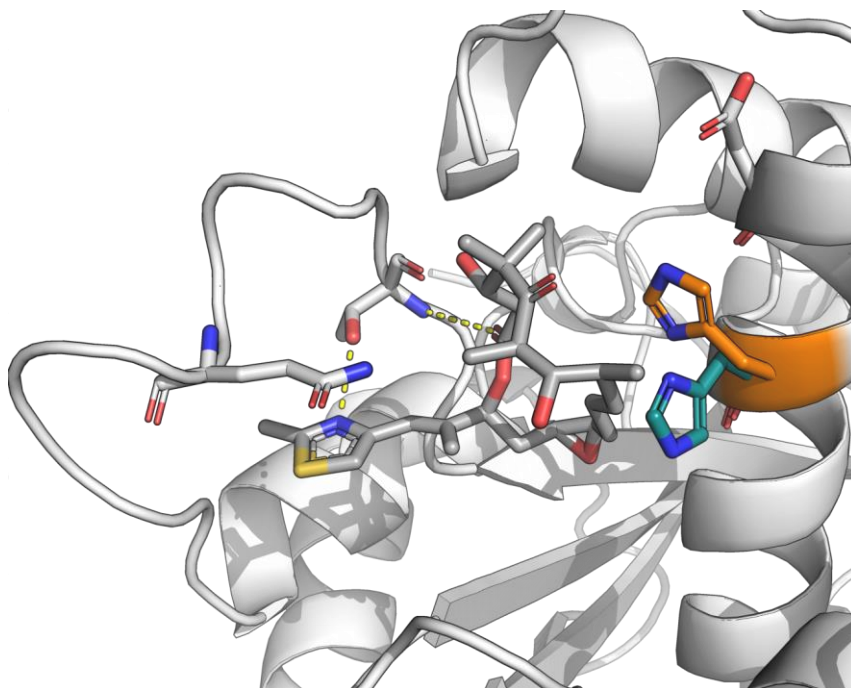


Figure 36 – The orientation of the His227 is important for the correct positioning of epothilone A macrolidic core. His227 is colored in orange when its position, as observed in many MD-Binding runs, does not allow the proper positioning of epothilone A. In green, as reported in the crystallographic structure (PDB ID: 4I50).

#### 4.3.3.3. Scoring

Staying on the traditional docking analogy, the MD-Binding approach can be seen as the searching algorithm. Therefore, it accounts only for the sampling part of the task and needs a scoring function to discriminate among all the obtained poses. In principle, one could monitor all the interactions that are formed or broken during the course of the simulations, rebuild the ligand's path to binding and evaluate the free energy associated to it. However, given the number of obtained poses (and the relative paths), this would have been a cumbersome task. Our aim was that of using a docking-like scoring function, fast and easy to handle, but enough accurate to weight the different poses.

We put our attention to the last frames of the MD-Binding simulations, where the bias has been switched off and the simulations reached a plain-MD-like status. We extracted the protein and the ligand only, and tested the obtained binding mode using a classical docking scoring function, the GLIDE SP one, to test its ability of selecting among all the poses, the one similar to the X-ray crystal structure. In order to remove any clash between the protein

and the ligands, we select to minimize the complex. Moreover, we removed all the waters from the computation because, for the way the GLIDE SP scoring function works, they all would have been considered as positive interactions, even that waters that are not bridging interactions with the protein. The results for the set of simulations run with the reference setup 12 (Table 4) are shown in Table 10 where it is possible to note that, according to the GLIDE SP score, the best pose has an RMSD value of 7.43 Å in respect to the X-ray crystal structure.

Run ID	RMSD against X-Ray Structure (Å)	GLIDE SP Score
1	7.4408	-4.351
2	9.2039	-3.465
3	9.8533	-5.800
4	7.4249	-4.443
5	1.6150	-6.591
6	7.0734	-4.624
7	8.5892	-3.587
8	7.0393	-6.486
9	8.5903	-4.913
10	8.7936	-2.988
11	5.0268	-5.193
12	7.4327	-7.783
13	8.9014	-4.942
14	9.8397	-6.338
15	8.9710	-2.919
16	11.8945	-3.244

Table 10 – GLIDE score for the 16 different poses obtained by means of MD-Binding (Table 4 - reference setup 12) and the relative RMSD to the X-ray structure (PDB ID: 4I50). The best-scoring pose is highlighted in red. The one that is closest to the X-ray structure in green.

For this reason, we decided to use a more accurate scoring function, the Prime MM/GBSA approach of the Schrödinger Suite 2016-2. Also for this method, we select to optimize all the residues at 6 Å from the ligand in order to avoid steric clashes between the ligand and the protein and reduce possible angle deviation from the equilibrium parameters.

Even though is known that MM-GBSA absolute energy values have no physical meaning, in this case, the best-ranked pose is the one that correctly reproduced the crystallographic binding mode (Table 11).

Rank	Run ID	RMSD against X-Ray Structure (Å)	$\Delta G_{\text{bind}}$ (kcal/mol)
1	5	1.6150	<b>-72.07</b>
2	14	9.8397	<b>-71.09</b>
3	12	7.4327	<b>-68.74</b>
4	11	5.0268	<b>-62.30</b>
5	4	7.4249	<b>-57.27</b>
6	9	8.5903	<b>-56.02</b>
7	6	7.0734	<b>-55.18</b>
8	8	7.0393	<b>-54.52</b>
9	3	9.8533	<b>-52.36</b>
10	1	7.4408	<b>-51.08</b>
11	13	8.9014	<b>-49.29</b>
12	2	9.2039	<b>-44.45</b>
13	7	8.5892	<b>-42.54</b>
14	16	11.8945	<b>-41.99</b>
15	10	8.7936	<b>-41.09</b>
16	15	8.9710	<b>-39.88</b>

Table 11 –  $\Delta G_{\text{bind}}$  (kcal/mol) calculated using Prime MM/GBSA method for the 16 different poses obtained by means of MD-Binding (Table 4 - reference setup 12) and the relative RMSD to the X-ray structure (PDB ID: 4I50).

We repeated the analysis for other two set of simulations run with the same setup (Table 4 - reference 12) and found results in agreement to those above-reported.

## 4.4. Dynamic Docking of Discodermolide and Taxol

Once validated our protocol with the epothilone A case, we decided to repeat our simulations with other two known binders of the taxane-site, discodermolide and taxol. For these cases, in order to speed-up the sampling phase, instead of using the whole  $\alpha\beta$ -tubulin heterodimer we used only a  $\beta$ -tubulin monomer. Essentially, in the short time of an MD-Binding run (10-30 ns), it is unlikely to observe large conformational changes in the protein structure. Nonetheless, we re-run a set of MD-Binding simulations for epothilone A and with the new setup (Table 4 - reference setup 13), we were able to reproduce the crystallographic binding mode and to correctly rank it with MM/GBSA.

### 4.4.1. Discodermolide

At the beginning of this study, conversely to the epothilone A case, no X-ray structures of discodermolide were available in the protein data bank. Therefore, we proceeded in a similar way as for traditional docking studies. Starting from the 2D structure (see Table 3), we generated an initial guess of the tridimensional conformation using the LigPrep utility of Schrödinger Suite 2016-4. Due to the high number of degrees of freedom of the molecule, we then performed an accurate conformational search with the MacroModel software using the same parameters as reported in ref [171,172] that resulted to produce results in agreement with NMR studies for the DDM conformation in solution. We further optimized the resulting lowest energy structure at the ab-initio HF/6-31G\* level theory with the Gaussian code and derived the RESP partial charges derived as for the epothilone A molecule.

Despite discodermolide binds at the same site of taxol, there were enough evidences by NMR [182], photo labeling [183], and HDX MS (Hydrogen-Deuterium Exchange coupled to Mass-Spectroscopy) [184] studies, that the two exert their activity through two different mechanisms, having different effects on the M-loop stability. After we started our simulations (Table 4 - reference setup 14), the X-ray structure of T2R-TTL-DDM complex came out (PDB ID: 5XLT) [185] and we had the possibility to check the crystallographic binding mode and compare it to the one of epothilone A and the one resulted by our simulations.

As expected, DDM does not make the same interactions as epothilone A. Specifically, the former does not interact with the Gln279, and thus it does not stabilize the M-loop to an  $\alpha$ -helix form. For this reason, there is no density associated to some of the residues of the loop in the X-ray crystal structure, similarly to what is reported for the apo form of tubulin,



PDB ID 4I55. Besides, DDM interacts with the S9-S10 loop, at the level of the backbone of Arg359 (Figure 37). There is also a water-mediated h-bond with the Ser230.

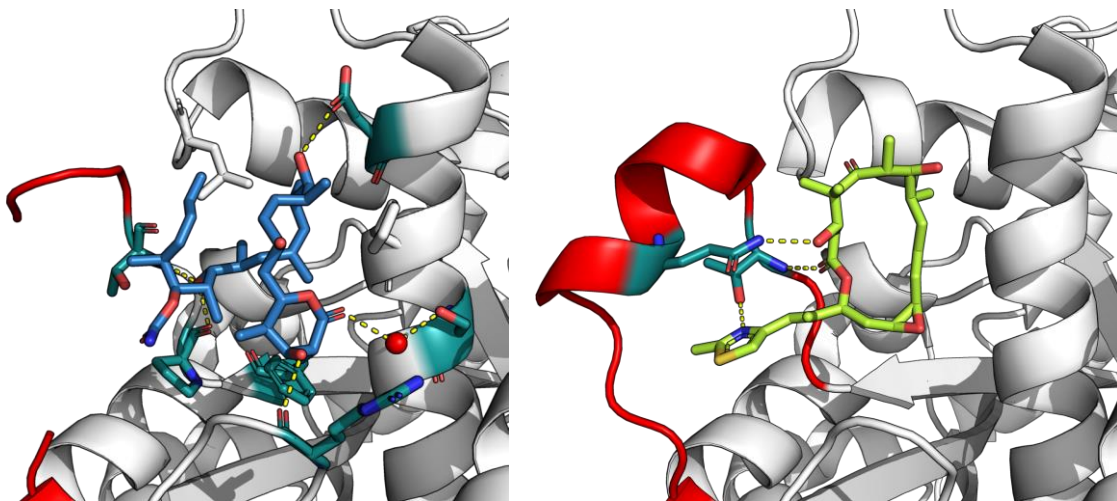


Figure 37 – Comparison between the binding mode of DDM (on the left) and EP (on the right) as reported in their respective X-ray structures. The highly flexible M-loop, not interacting with DDM, is not resolved.

Thanks to the available structure, we were able to monitor the RMSD of the ligand and of the binding site residues during the course of our simulations. Looking at the graph (Figure 38) is possible to note that for a set of 20 simulations, we recovered the X-ray binding mode three times.

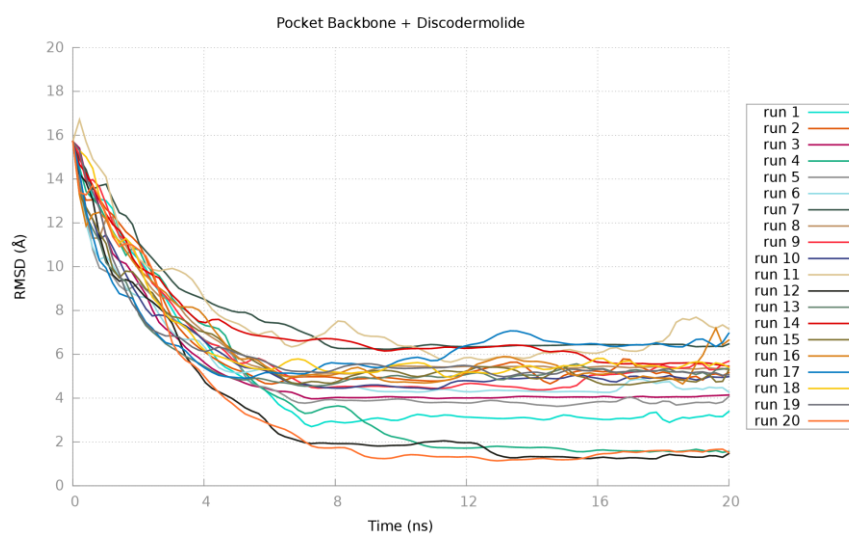


Figure 38 - Calculated RMSD of DDM and pocket backbone to the X-ray structure (PDB ID: 5LXT) plotted as a function of time.

We then looked carefully at the final poses and noticed a remarkable result: not only all the main interaction with the protein were recovered but also the water-mediated hydrogen bond with the Ser230 (Figure 39).

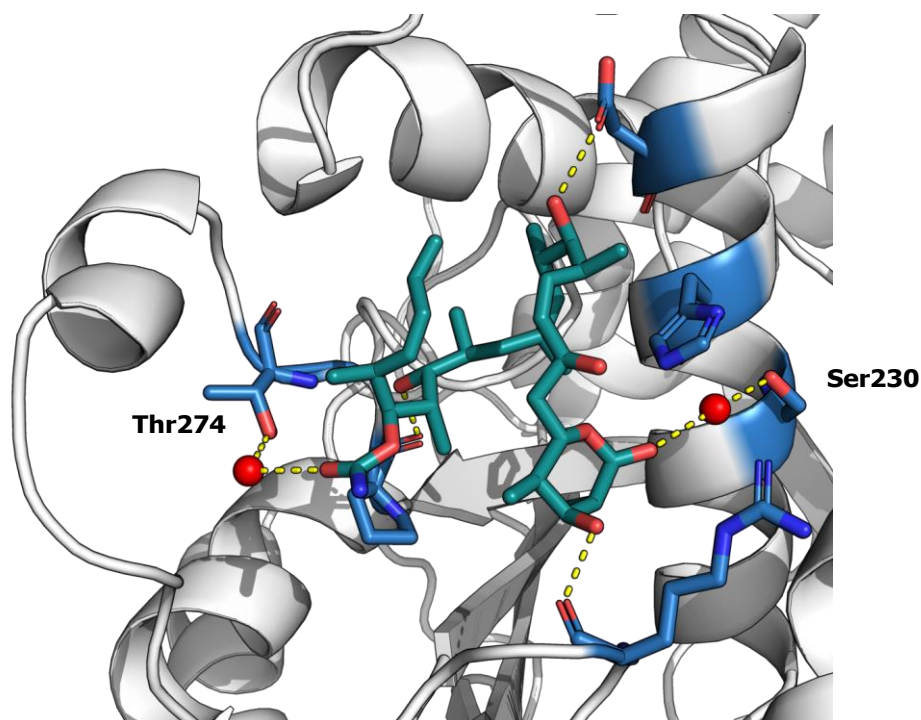


Figure 39 – The correctly reproduced binding mode of DDM with our MD-Binding protocol (Reference setup 14 of Table 4, Run ID 4). For comparison, see Figure 37 (left). The induced  $\alpha$ -helix is still well folded.

Besides, our simulations provided more information than the static view of the crystal structure. We noticed a fluctuation in the hydrogen bond between the nitrogen of the backbone of Thr274 and carbonyl group of DDM. This was due to the formation of a weak interaction between the carbamate group of DDM and the side-chain of Thr274 (that is not present in the X-ray structure).

Moving to the scoring phase, Prime MM/GBSA resulted useful also in this case. It showed to be capable of ranking at the top places all the three simulations whose binding mode was closer (in terms of RMSD) to the X-ray crystal structure (PDB ID: 5LXT). The results are summarized in Table 12.

Rank	Run ID	RMSD against X-Ray Structure (Å)	$\Delta G_{\text{bind}}$ (kcal/mol)
1	20	5.17	<b>-84.94</b>
2	12	8.28	<b>-76.77</b>
3	4	6.21	<b>-69.64</b>
4	5	1.20	<b>-64.34</b>
5	6	6.01	<b>-56.78</b>
6	9	6.74	<b>-56.75</b>
7	14	8.97	<b>-55.48</b>
8	19	7.55	<b>-52.57</b>
9	8	9.08	<b>-52.35</b>
10	1	7.67	<b>-52.29</b>
11	3	11.4	<b>-50.41</b>
12	18	1.95	<b>-48.61</b>
13	11	8.52	<b>-46.31</b>
14	17	6.93	<b>-43.15</b>
15	10	8.53	<b>-43.07</b>
16	7	10.4	<b>-42.06</b>
17	13	11.2	<b>-38.06</b>
18	16	8.60	<b>-36.26</b>
19	2	7.94	<b>-30.81</b>
20	15	1.77	<b>-30.45</b>

Table 12 –  $\Delta G_{\text{bind}}$  (kcal/mol) calculated using Prime MM/GBSA method for the 20 different poses obtained by means of MD-Binding (Table 4 - reference setup 14) and the relative RMSD to the X-ray structure (PDB ID: 5LXT).

#### 4.4.2. Taxol

Lastly, we performed a number of MD-Binding simulations for taxol, starting from the optimization of the 2D structure analogously to what we did for discodermolide. Despite the low resolution of the available crystal structure (PDB ID: 1JFF – 3.5 Å), the binding mode proposed by the authors is in agreement with the results obtained by the photoaffinity labeling experiments of a series of taxol analogs [186]. Therefore, we decided to compare the outcome our simulations to the binding mode proposed in the 2001 by electron crystallography of Zn-induced tubulin sheets (see Paragraph 2.5.1 and Figure 10). Specifically, we conducted a series of 30-ns long simulations using the same setup as for EP and DDM (Table 4 - reference setup 15). For the comparison, we considered only those simulations in which taxol made contact with the switching-off residue Pro272, hence a plain MD-like status has been recovered (for visual explanation, see Figure 40).

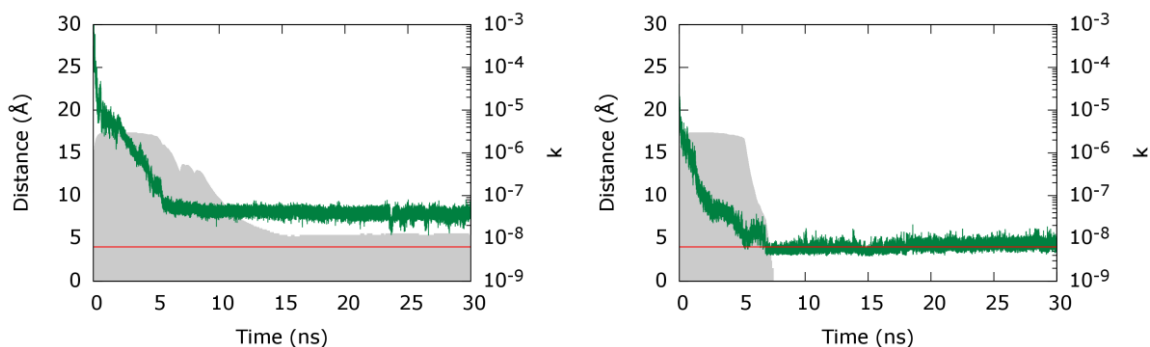


Figure 40 – Comparison of bias and distance evolution during the course of TAX MD-Binding simulations. On the left, a simulation in which TAX did not make contact with the switching-off residue Pro272; on the right, the bias is completely switched off at about 7 ns, and the simulation reached a plain MD-like status.

Unfortunately, none of the simulations recovered the crystallographic binding mode, thus we did not proceed to the scoring phase.

However, visual inspecting our simulations, we noticed that in most of them the bias switched off because of the contacts made by one of the lateral phenyl groups, rather than by the baccatin main core (Figure 41).

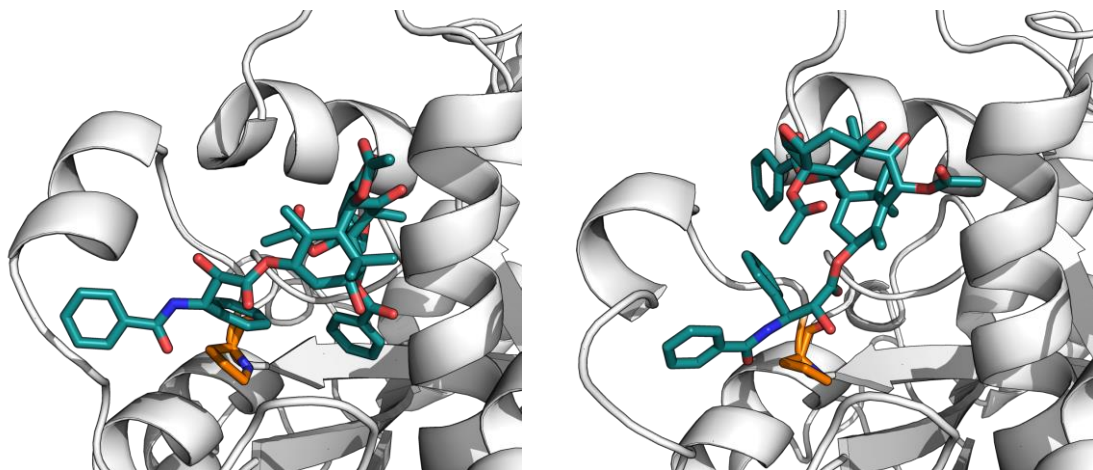
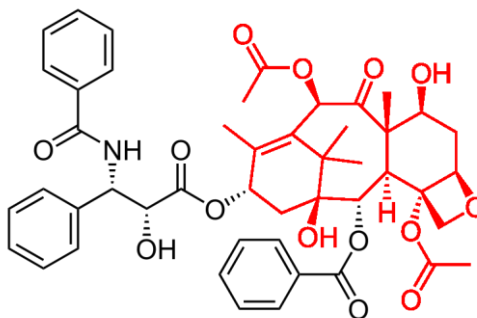


Figure 41 – The taxol obtained discordant binding modes. The selected switching-off residue Pro272 is shown in orange.

In light of these results, we hypothesize that a different selection for the taxol group should have been done. In particular, instead of spreading the virtual charges across all the

taxol atoms, we should have selected only those forming the baccatin core (Figure 42). In analogy to traditional docking techniques, this could resemble, in some way, the constraining of a particular set of atoms, a procedure particularly useful when dealing with a congeneric series of compounds.



---

Figure 42 – The taxol baccatin core highlighted in red.

In our case, it could have driven the accommodation of the main core and, after the bias switching-off, to the subsequent arrangement of the lateral chains.

Nonetheless, further simulations and analysis are required.

## 5. Conclusions and Perspectives

Even though there are no doubts about the importance of the role played by docking simulations in discovering and developing novel compounds of pharmaceutical interest, to make sense of often-discordant experimental results and to gain a better understanding about the driving forces of a binding event, a paradigm shift to a dynamical approach is mandatory. Proteins are inherently flexible entities and, especially for those cases in which large conformational rearrangements are observed upon binding, the static picture provided by traditional docking techniques can't elucidate the mechanisms at the basis of the binding process. In line with this reasoning, molecular dynamics and the related enhanced sampling methods are everyday gaining more importance and, in the last few years, several computational approaches have been developed in order to exploit their potential in drug discovery and development programs.

In this dissertation, we made use of one of these state-of-the-art strategies, the MD-Binding technique, to devise a protocol for the dynamic docking of taxane-site ligands to tubulin, a protein that by its own nature can exist in three major forms: heterodimers, curved oligomers, and straight assembled microtubules. We showed that molecular dynamics methods, by taking into account protein flexibility of both backbone and side-chains, have the capabilities to overcome many of the limitations of which traditional docking techniques suffer.

After an initial plain MD simulation, thanks to which we gathered valuable information regarding the behavior of the binding pocket and, specifically, of its highly flexible M-loop, we validated our MD-Binding protocol against the remarkable case of epothilone A. Its binding mode, in fact, remained elusive until recently, when the high-res X-ray crystal structure of the drug in complex with a curved  $\alpha/\beta$ -tubulin heterodimer was determined and showed clear differences from the one proposed using EC data of Zn-induced tubulin sheets. Our protocol was able to discriminate between the two selecting the correct one. Besides, by comparing it with a traditional software, we showed how the MD-Binding approach relieves the user by the arduous work of choosing one or more suitable protein conformations against which to perform the docking simulations.

We had similar results for another ligand, discodermolide, but not for the well-known microtubule-stabilizing agent taxol. For the latter, although there are no high-res X-ray crystal structures to compare with and verify our results, our protocol was not able to select

any relevant binding mode for the drug. We, therefore, hypothesized that a less generic approach must be evaluated when dealing with a ligand of its size but further simulations and analysis are needed.

In conclusion, our study showed the feasibility and the potential of the MD-Binding method in correctly reproducing the binding modes of two highly flexible ligands, epothilone A and discodermolide, in the as much flexible binding site of tubulin at a fraction of the time that is generally needed by plain MD simulations. However, the scoring phase represents still an open issue for every computational method, and further advancements in the field are required. Nonetheless, we believe that, with the advent of more powerful hardware architectures and of more accurate and efficient algorithms, dynamic docking will take the place of traditional methods at any level of drug discovery programs.



# APPENDIX: Residue Numbering Correspondence

	Helix H1																													
	MET	ARG	GLU	ILE	VAL	HIS	ILE	GLN	ALA	GLY	GLN	CYS	GLY	ASN	GLN	ILE	GLY	ALA	LYS	PHE	TRP	GLU	VAL	ILE	SER	ASP	GLU	HIS	GLY	ILE
Lowe et al.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
UniProtKB	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Heterodimer	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470
	ASP	PRO	THR	GLY	SER	TYR	HIS	GLY	ASP	SER	ASP	LEU	GLN	LEU	GLU	ARG	ILE	ASN	VAL	TYR	TYR	ASN	GLU	ALA	THR	GLY	ASN	LYS	TYR	VAL
Lowe et al.	31	32	33	34	35	36	37	38	39	40	41	42	43	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
UniProtKB	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Heterodimer	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500
	PRO	ARG	ALA	ILE	LEU	VAL	ASP	LEU	GLU	PRO	GLY	THR	MET	ASP	SER	VAL	ARG	SER	GLY	PRO	PHE	GLY	GLN	ILE	PHE	ARG	PRO	ASP	ASN	PHE
Lowe et al.	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
UniProtKB	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Heterodimer	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530
	VAL	PHE	GLY	GLN	SER	GLY	ALA	GLY	ASN	ASN	TRP	ALA	LYS	GLY	HIS	TYR	THR	GLU	GLY	ALA	GLU	LEU	VAL	ASP	SER	VAL	LEU	ASP	VAL	VAL
Lowe et al.	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122
UniProtKB	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
Heterodimer	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560
	ARG	LYS	GLU	SER	GLU	SER	CYS	ASP	CYS	LEU	GLN	GLY	PHE	GLN	LEU	THR	HIS	SER	LEU	GLY	GLY	GLY	THR	GLY	SER	GLY	MET	GLY	THR	LEU
Lowe et al.	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152
UniProtKB	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
Heterodimer	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590
	LEU	ILE	SER	LYS	ILE	ARG	GLU	GLU	TYR	PRO	ASP	ARG	ILE	MET	ASN	THR	PHE	SER	VAL	MET	PRO	SER	PRO	LYS	VAL	SER	ASP	THR	VAL	VAL
Lowe et al.	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182
UniProtKB	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
Heterodimer	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620
	Helix H6																													
	GLU	PRO	TYR	ASN	ALA	THR	LEU	SER	VAL	HIS	GLN	LEU	VAL	GLU	ASN	THR	ASP	GLU	THR	TYR	CYS	ILE	ASP	ASN	GLU	ALA	LEU	TYR	ASP	ILE
Lowe et al.	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212
UniProtKB	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
Heterodimer	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650
	Helix H7																													
	CYS	PHE	ARG	THR	LEU	LYS	LEU	THR	THR	PRO	THR	TYR	GLY	ASP	LEU	ASN	HIS	LEU	VAL	SER	ALA	THR	MET	SER	GLY	VAL	THR	THR	CYS	LEU
Lowe et al.	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242
UniProtKB	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
Heterodimer	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680
	Strand S7																													
	ARG	PHE	PRO	GLY	GLN	LEU	ASN	ALA	ASP	LEU	ARG	LYS	LEU	ALA	VAL	ASN	MET	VAL	PRO	PHE	PRO	ARG	LEU	HIS	PHE	PHE	MET	PRO	ARG	PHE
Lowe et al.	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272
UniProtKB	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270
Heterodimer	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710
	M-Loop																													
	ALA	PRO	LEU	THR	SER	ARG	GLY	SER	GLN	GLN	TYR	ARG	ALA	LEU	THR	VAL	PRO	GLU	LEU	THR	GLN	GLN	MET	PHE	ASP	SER	LYS	ASN	MET	MET
Lowe et al.	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302
UniProtKB	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
Heterodimer	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740
	ALA	ALA	CYS	ASP	PRO	ARG	HIS	GLY	ARG	TYR	LEU	THR	VAL	ALA	ALA	ILE	PHE	ARG	GLY	ARG	MET	SER	MET	LYS	GLU	VAL	ASP	GLU	GLN	MET
Lowe et al.	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332
UniProtKB	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330
Heterodimer	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770
	Loop S9-S10																													
	LEU	ASN	VAL	GLN	ASN	LYS	ASN	SER	SER	TYR	PHE	VAL	GLU	TRP	ILE	PRO	ASN	ASN	VAL	LYS	THR	ALA	VAL	CYS	ASP	ILE	PRO	PRO	ARG	GLY
Lowe et al.	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362
UniProtKB	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360
Heterodimer	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800
	Loop S11-S12																													
	LEU	LYS	MET	SER	ALA	THR	PHE	ILE	GLY	ASN	SER	THR	ALA	ILE	GLN	GLY	PHE	THR	GLY	ARG	ILE</									



gene TUBB2B. This residue numbering was used throughout the dissertation when referring to the monomeric system. In the  $\alpha\beta$ -tubulin heterodimeric system, the numbering is shifted by 440, which is the number of residues of the  $\alpha$ -tubulin subunit.

Lowe *et al.* introduced two gaps (at the level of residues that colored in red in the Table) due to the alignment of  $\alpha$  and  $\beta$  tubulin subunits.  $\alpha$ -tubulin, this is more evident when considering the second gap (Pro360-Arg369), has 9 more residues in the corresponding loop.

The three protein domains are colored differently for clarity reasons: the N-terminal domain in light yellow, the drug-binding domain in light blue, and the C-terminal domain in light magenta.

Secondary structures that create the taxane-site are highlighted.

## REFERENCES

1. Kapetanovic, I. M. Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach. *Chem. Biol. Interact.* **2008**, *171*, 165–76, doi:10.1016/j.cbi.2006.12.006.
2. Cumming, J. G.; Davis, A. M.; Muresan, S.; Haeberlein, M.; Chen, H. Chemical predictive modelling to improve compound quality. *Nat. Rev. Drug Discov.* **2013**, *12*, 948–962, doi:10.1038/nrd4128.
3. Carnero, A. High throughput screening in drug discovery. *Clin. Transl. Oncol.* **2006**, *8*, 482–90.
4. McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494–502.
5. Mullard, A. The drug-maker's guide to the galaxy. *Nature* **2017**, *549*, 445–447, doi:10.1038/549445a.
6. Talele, T.; Khedkar, S.; Rigby, A. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* **2010**, *10*, 127–141, doi:10.2174/156802610790232251.
7. Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–8, doi:10.1126/science.1096361.
8. Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* **2014**, *14*, 1923–38, doi:10.2174/1568026614666140929124445.
9. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–88.
10. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242, doi:10.1093/nar/28.1.235.
11. Matter, H.; Sotriffer, C. *Applications and Success Stories in Virtual Screening*; Wiley-VCH Verlag GmbH & Co. KGaA, 2011; ISBN 9783527633326.
12. Carlson, H. A. Protein flexibility and drug design: how to hit a moving target. *Curr.*

- Opin. Chem. Biol.* **2002**, *6*, 447–452, doi:10.1016/S1367-5931(02)00341-1.
13. Copeland, R. A. The drug–target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* **2015**, *15*, 87–95, doi:10.1038/nrd.2015.18.
  14. Bernetti, M.; Cavalli, A.; Mollica, L. Protein–ligand (un)binding kinetics as a new paradigm for drug discovery at the crossroad between experiments and modelling. *Med. Chem. Commun.* **2017**, *8*, 534–550, doi:10.1039/C6MD00581K.
  15. Lill, M. A. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry* **2011**, *50*, 6157–69, doi:10.1021/bi2004558.
  16. Feixas, F.; Lindert, S.; Sinko, W.; McCammon, J. A. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys. Chem.* **2014**, *186*, 31–45, doi:10.1016/j.bpc.2013.10.007.
  17. Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* **2003**, *2*, 527–541, doi:10.1038/nrd1129.
  18. Bottegoni, G. Protein-ligand docking. *Front. Biosci. (Landmark Ed.)* **2011**, *16*, 2289–2306, doi:http://dx.doi.org/10.2741/3854.
  19. Koshland, D. E.; Jr. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98–104.
  20. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598–603.
  21. Monod, J.; Wyman, J.; Changeux, J. P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88–118, doi:10.1016/S0022-2836(65)80285-6.
  22. Buonfiglio, R.; Recanatini, M.; Masetti, M. Protein Flexibility in Drug Discovery: From Theory to Computation. *ChemMedChem* **2015**, *10*, 1141–1148, doi:10.1002/cmdc.201500086.
  23. Sherman, W.; Beard, H. S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* **2006**, *67*, 83–84.
  24. Rueda, M.; Bottegoni, G.; Abagyan, R. Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *J. Chem. Inf. Model.* **2010**, *50*, 186–193, doi:10.1021/ci9003943.

25. Bottegoni, G.; Rocchia, W.; Rueda, M.; Abagyan, R.; Cavalli, A. Systematic Exploitation of Multiple Receptor Conformations for Virtual Ligand Screening. *PLoS One* **2011**, *6*, e18845, doi:10.1371/journal.pone.0018845.
26. Craig, I. R.; Essex, J. W.; Spiegel, K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 511–524, doi:10.1021/ci900407c.
27. Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. Receptor–ligand molecular docking. *Biophys. Rev.* **2014**, *6*, 75–87, doi:10.1007/s12551-013-0130-2.
28. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047–69, doi:10.1016/S0006-3495(97)78756-3.
29. Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55*, 475–482, doi:10.1021/ci500731a.
30. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061, doi:10.1021/acs.jmedchem.5b01684.
31. Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633, doi:10.1021/ja0260162.
32. Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided. Mol. Des.* **2008**, *22*, 693–705, doi:10.1007/s10822-007-9159-2.
33. De Paris, R.; Quevedo, C. V.; Ruiz, D. D.; Norberto De Souza, O.; Barros, R. C. Clustering molecular dynamics trajectories for optimizing docking experiments. *Comput. Intell. Neurosci.* **2015**, *2015*, 916240, doi:10.1155/2015/916240.
34. Offutt, T. L.; Swift, R. V.; Amaro, R. E. Enhancing Virtual Screening Performance of Protein Kinases with Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2016**, *56*, 1923–1935, doi:10.1021/acs.jcim.6b00261.
35. Osguthorpe, D. J.; Sherman, W.; Hagler, A. T. Generation of Receptor Structural Ensembles for Virtual Screening Using Binding Site Shape Analysis and Clustering.

- Chem. Biol. Drug Des.* **2012**, *80*, 182–193, doi:10.1111/j.1747-0285.2012.01396.x.
36. Cavalli, A.; Bottegoni, G.; Raco, C.; De Vivo, M.; Recanatini, M. A computational study of the binding of propidium to the peripheral anionic site of human acetylcholinesterase. *J. Med. Chem.* **2004**, *47*, 3991–3999, doi:10.1021/jm040787u.
  37. Perdih, A.; Hrast, M.; Pureber, K.; Barreteau, H.; Grdadolnik, S. G.; Kocjan, D.; Gobec, S.; Solmajer, T.; Wolber, G. Furan-based benzene mono- and dicarboxylic acid derivatives as multiple inhibitors of the bacterial Mur ligases (MurC–MurF): experimental and computational characterization. *J. Comput. Aided. Mol. Des.* **2015**, *29*, 541–560, doi:10.1007/s10822-015-9843-6.
  38. Sakano, T.; Mahamood, M. I.; Yamashita, T.; Fujitani, H. Molecular dynamics analysis to evaluate docking pose prediction. *Biophys. physicobiology* **2016**, *13*, 181–194, doi:10.2142/biophysico.13.0\_181.
  39. Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531–68, doi:10.1002/med.20067.
  40. Decherchi, S.; Masetti, M.; Vyalov, I.; Rocchia, W. Implicit solvent methods for free energy estimation. *Eur. J. Med. Chem.* **2015**, *91*, 27–42.
  41. Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426, doi:10.1063/1.1740409.
  42. Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313, doi:10.1063/1.1749657.
  43. De Vivo, M.; Cavalli, A. Recent advances in dynamic docking for drug discovery. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7*, e1320, doi:10.1002/wcms.1320.
  44. Gioia, D.; Bertazzo, M.; Recanatini, M.; Masetti, M.; Cavalli, A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. *Molecules* **2017**, *22*, 2029, doi:10.3390/molecules22112029.
  45. Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58–65, doi:10.1016/j.sbi.2012.11.002.
  46. Noé, F. Beating the millisecond barrier in molecular dynamics simulations. *Biophys.*

- J.* **2015**, *108*, 228–229, doi:10.1016/j.bpj.2014.11.3477.
47. Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2013**, *16*, 163–199, doi:10.3390/e16010163.
  48. Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181–9183, doi:10.1021/ja500545u.
  49. Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **2011**, *108*, 10184–10189, doi:10.1073/pnas.1103547108.
  50. Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci.* **2011**, *108*, 13118–13123, doi:10.1073/pnas.1104614108.
  51. Kruse, A. C.; Hu, J.; Pan, A. C.; Arlow, D. H.; Rosenbaum, D. M.; Rosemond, E.; Green, H. F.; Liu, T.; Chae, P. S.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; Wess, J.; Kobilka, B. K. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **2012**, *482*, 552–556, doi:10.1038/nature10867.
  52. Decherchi, S.; Berteotti, A.; Bottegoni, G.; Rocchia, W.; Cavalli, A. The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat. Commun.* **2015**, *6*, 6155, doi:10.1038/ncomms7155.
  53. Ferruz, N.; Harvey, M. J.; Mestres, J.; De Fabritiis, G. Insights from Fragment Hit Binding Assays by Molecular Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2200–2205, doi:10.1021/acs.jcim.5b00453.
  54. Bisignano, P.; Doerr, S.; Harvey, M. J.; Favia, A. D.; Cavalli, A.; De Fabritiis, G. Kinetic characterization of fragment binding in AmpC  $\beta$ -lactamase by high-throughput molecular simulations. *J. Chem. Inf. Model.* **2014**, *54*, 362–366, doi:10.1021/ci4006063.
  55. Dror, R. O.; Green, H. F.; Valant, C.; Borhani, D. W.; Valcourt, J. R.; Pan, A. C.; Arlow, D. H.; Canals, M.; Lane, J. R.; Rahmani, R.; Baell, J. B.; Sexton, P. M.;

- Christopoulos, A.; Shaw, D. E. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **2013**, 2–9, doi:10.1038/nature12595.
56. Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 2010, 52, 99–105.
  57. Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* 2014, 25, 135–144.
  58. Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, 6, 7653, doi:10.1038/ncomms8653.
  59. Doerr, S.; De Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J. Chem. Theory Comput.* **2014**, 10, 2064–2069, doi:10.1021/ct400919u.
  60. Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **2016**, 12, 1845–1852, doi:10.1021/acs.jctc.6b00049.
  61. Ferruz, N.; Tresadern, G.; Pineda-Lucena, A.; De Fabritiis, G. Multibody cofactor and substrate molecular recognition in the myo-inositol monophosphatase enzyme. *Sci. Rep.* **2016**, 6, 30275, doi:10.1038/srep30275.
  62. Stanley, N.; Pardo, L.; Fabritiis, G. De The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Sci. Rep.* **2016**, 6, 22639, doi:10.1038/srep22639.
  63. Sabbadin, D.; Moro, S. Supervised Molecular Dynamics (SuMD) as a Helpful Tool To Depict GPCR–Ligand Recognition Pathway in a Nanosecond Time Scale. *J. Chem. Inf. Model.* **2014**, 54, 372–376, doi:10.1021/ci400766b.
  64. Sabbadin, D.; Cianchetta, A.; Deganutti, G.; Cuzzolin, A.; Moro, S. Exploring the recognition pathway at the human A<sub>2A</sub> adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations. *Med. Chem. Commun.* **2015**, 6, 1081–1085, doi:10.1039/C5MD00016E.
  65. Cuzzolin, A.; Sturlese, M.; Deganutti, G.; Salmaso, V.; Sabbadin, D.; Cianchetta, A.; Moro, S. Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J. Chem. Inf. Model.* **2016**, 56,

687–705, doi:10.1021/acs.jcim.5b00702.

66. Zeller, F.; Luitz, M. P.; Bomblies, R.; Zacharias, M. Multiscale Simulation of Receptor–Drug Association Kinetics: Application to Neuraminidase Inhibitors. *J. Chem. Theory Comput.* **2017**, *acs.jctc.7b00631*, doi:10.1021/acs.jctc.7b00631.
67. Ermak, D. L.; McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352–1360, doi:10.1063/1.436761.
68. Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–6, doi:10.1073/pnas.202427399.
69. Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **1996**, *271*, 997–9.
70. Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199, doi:10.1016/0021-9991(77)90121-8.
71. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151, doi:10.1016/S0009-2614(99)01123-9.
72. Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929, doi:10.1063/1.1755656.
73. Mark, A. E.; van Gunsteren, W. F.; Berendsen, H. J. C. Calculation of relative free energy via indirect pathways. *J. Chem. Phys.* **1991**, *94*, 3808–3816, doi:10.1063/1.459753.
74. Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *J. Phys. Chem. B* **1997**, *101*, 817–824, doi:doi: 10.1021/jp962142e.
75. Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607, doi:10.1021/ja0445950.
76. Provasi, D.; Bortolato, A.; Filizola, M. Exploring molecular mechanisms of ligand recognition by opioid receptors with metadynamics. *Biochemistry* **2009**, *48*, 10020–10029, doi:10.1021/bi901494n.



77. Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci.* **2013**, *110*, 6358–6363, doi:10.1073/pnas.1303186110.
78. Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports Prog. Phys.* **2008**, *71*, 126601, doi:10.1088/0034-4885/71/12/126601.
79. Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553–4559, doi:10.1021/jp067873l.
80. Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. *J. Am. Chem. Soc.* **2009**, *131*, 11811–11818, doi:10.1021/ja903045y.
81. Soderhjelm, P.; Tribello, G. A.; Parrinello, M. Locating binding poses in protein-ligand systems using reconnaissance metadynamics. *Proc. Natl. Acad. Sci.* **2012**, *109*, 5170–5175, doi:10.1073/pnas.1201940109.
82. Yoshida, K.; Yamaguchi, T.; Okamoto, Y. Replica-exchange molecular dynamics simulation of small peptide in water and in ethanol. *Chem. Phys. Lett.* **2005**, *412*, 280–284, doi:10.1016/j.cplett.2005.06.114.
83. Ostermeir, K.; Zacharias, M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochim. Biophys. Acta* **2013**, *1834*, 847–853, doi:10.1016/j.bbapap.2012.12.016.
84. Luitz, M. P.; Zacharias, M. Protein-ligand docking using Hamiltonian replica exchange simulations with soft core potentials. *J. Chem. Inf. Model.* **2014**, *54*, 1669–1675, doi:10.1021/ci500296f.
85. Kappel, K.; Miao, Y.; McCammon, J. A. Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. *Q. Rev. Biophys.* **2015**, *48*, 479–487, doi:10.1017/S0033583515000153.
86. Kamiya, N.; Yonezawa, Y.; Nakamura, H.; Higo, J. Protein-inhibitor flexible docking by a multicanonical sampling: Native complex structure with the lowest free energy and a free-energy barrier distinguishing the native complex from the others. *Proteins Struct. Funct. Bioinforma.* **2007**, *70*, 41–53, doi:10.1002/prot.21409.
87. Bekker, G.-J.; Kamiya, N.; Araki, M.; Fukuda, I.; Okuno, Y.; Nakamura, H. Accurate

- Prediction of Complex Structure and Affinity for a Flexible Protein Receptor and Its Inhibitor. *J. Chem. Theory Comput.* **2017**, *13*, 2389–2399, doi:10.1021/acs.jctc.6b01127.
88. Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362, doi:10.1080/00268976.2013.813594.
  89. Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 2000, *10*, 153–159.
  90. Macchiarulo, A.; Costantino, G.; Sbaglia, R.; Aiello, S.; Meniconi, M.; Pellicciari, R. The role of electrostatic interaction in the molecular recognition of selective agonists to metabotropic glutamate receptors. *Proteins Struct. Funct. Bioinforma.* **2003**, *50*, 609–619, doi:10.1002/prot.10301.
  91. Li, M.; Hoeck, C.; Schoffelen, S.; Gotfredsen, C. H.; Meldal, M. Specific Electrostatic Molecular Recognition in Water. *Chem. - A Eur. J.* **2016**, *22*, 7206–7214, doi:10.1002/chem.201600231.
  92. Rocchia, W.; Cavalli, A. Studying molecular interaction via enhanced molecular dynamics simulations. *WO 2016024194 A1* 2016.
  93. Spitaleri, A.; Decherchi, S.; Cavalli, A.; Rocchia, W. Fast Dynamic Docking Guided by Adaptive Electrostatic Bias: The MD-Binding Approach. *J. Chem. Theory Comput.* **2018**, *14*, 1727–1736, doi:10.1021/acs.jctc.7b01088.
  94. Decherchi, S.; Bottegoni, G.; Spitaleri, A.; Rocchia, W.; Cavalli, A. BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery. *J. Chem. Inf. Model.* **2018**, *58*, 219–224, doi:10.1021/acs.jcim.7b00680.
  95. Avila, J. Microtubule functions. *Life Sci.* **1992**, *50*, 327–334, doi:10.1016/0024-3205(92)90433-P.
  96. Caviston, J. P.; Holzbaur, E. L. F. Microtubule motors at the intersection of trafficking and transport. *Trends Cell Biol.* 2006, *16*, 530–537.
  97. Tilney, L. G.; Bryan, J.; Bush, D. J.; Fujiwara, K.; Mooseker, M. S.; Murphy, D. B.; Snyder, D. H. Microtubules: Evidence for 13 protofilaments. *J. Cell Biol.* **1973**, *59*, 267–275, doi:10.1083/jcb.59.2.267.

98. Chrétien, D.; Metoz, F.; Verde, F.; Karsenti, E.; Wade, R. H. Lattice defects in microtubules: protofilament numbers vary within individual microtubules. *J. Cell Biol.* **1992**, *117*, 1031–40, doi:10.1083/JCB.117.5.1031.
99. Pilhofer, M.; Ladinsky, M. S.; McDowall, A. W.; Petroni, G.; Jensen, G. J. Microtubules in Bacteria: Ancient Tubulins Build a Five-Protofilament Homolog of the Eukaryotic Cytoskeleton. *PLoS Biol.* **2011**, *9*, e1001213, doi:10.1371/journal.pbio.1001213.
100. Díaz, J. F.; Valpuesta, J. M.; Chacón, P.; Diakun, G.; Andreu, J. M. Changes in microtubule protofilament number induced by Taxol binding to an easily accessible site. Internal microtubule dynamics. *J. Biol. Chem.* **1998**, *273*, 33803–10, doi:10.1074/JBC.273.50.33803.
101. McIntosh, J. R.; Morpew, M. K.; Grissom, P. M.; Gilbert, S. P.; Hoenger, A. Lattice Structure of Cytoplasmic Microtubules in a Cultured Mammalian Cell. *J. Mol. Biol.* **2009**, *394*, 177–182, doi:10.1016/J.JMB.2009.09.033.
102. Sui, H.; Downing, K. H. Structural basis of interprotofilament interaction and lateral deformation of microtubules. *Structure* **2010**, *18*, 1022–1031, doi:10.1016/j.str.2010.05.010.
103. Zhou, J. *Microtubule protocols*; Humana Press, 2007; ISBN 1588296423.
104. Ferreira, J. G.; Pereira, A. L.; Maiato, H. Microtubule Plus-End Tracking Proteins and Their Roles in Cell Division. *Int. Rev. Cell Mol. Biol.* **2014**, *309*, 59–140, doi:10.1016/B978-0-12-800255-1.00002-8.
105. Sanchez, A. D.; Feldman, J. L. Microtubule-organizing centers: from the centrosome to non-centrosomal sites. *Curr. Opin. Cell Biol.* **2017**, *44*, 93–101, doi:10.1016/J.CEB.2016.09.003.
106. Kollman, J. M.; Merdes, A.; Mourey, L.; Agard, D. A. Microtubule nucleation by  $\gamma$ -tubulin complexes. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 709–721, doi:10.1038/nrm3209.
107. Roostalu, J.; Surrey, T. Microtubule nucleation: beyond the template. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 702–710, doi:10.1038/nrm.2017.75.
108. Voter, W. A.; Erickson, H. P. The kinetics of microtubule assembly. Evidence for a two-stage nucleation mechanism. *J. Biol. Chem.* **1984**, *259*, 10430–8.

109. Rodionov, V. I.; Borisy, G. G. Microtubule treadmilling in vivo. *Science* **1997**, *275*, 215–8, doi:10.1126/SCIENCE.275.5297.215.
110. Chen, W.; Zhang, D. Kinetochore fibre dynamics outside the context of the spindle during anaphase. *Nat. Cell Biol.* **2004**, *6*, 227–231, doi:10.1038/ncb1104.
111. Kirschner, M. W. Implications of treadmilling for the stability and polarity of actin and tubulin polymers in vivo. *J. Cell Biol.* **1980**, *86*, 330–4.
112. Mitchison, T.; Kirschner, M. Dynamic instability of microtubule growth. *Nature* **1984**, *312*, 237–242, doi:10.1038/312237a0.
113. Schulze, E.; Kirschner, M. New features of microtubule behaviour observed in vivo. *Nature* **1988**, *334*, 356–359, doi:10.1038/334356a0.
114. Sammak, P. J.; Borisy, G. G. Direct observation of microtubule dynamics in living cells. *Nature* **1988**, *332*, 724–726, doi:10.1038/332724a0.
115. Gardner, M. K.; Zanic, M.; Howard, J. Microtubule catastrophe and rescue. *Curr. Opin. Cell Biol.* **2013**, *25*, 14–22, doi:10.1016/J.CEB.2012.09.006.
116. Hyman, A. A.; Salser, S.; Drechsel, D. N.; Unwin, N.; Mitchison, T. J. Role of GTP hydrolysis in microtubule dynamics: information from a slowly hydrolyzable analogue, GMPCPP. *Mol. Biol. Cell* **1992**, *3*, 1155–67, doi:10.1091/MBC.3.10.1155.
117. Desai, A.; Mitchison, T. J. Microtubule Polymerization Dynamics. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 83–117, doi:10.1146/annurev.cellbio.13.1.83.
118. Nogales, E.; Wolf, S. G.; Downing, K. H. Structure of the  $\alpha\beta$  tubulin dimer by electron crystallography. *Nature* **1998**, *391*, 199–203, doi:10.1038/34465.
119. Löwe, J.; Li, H.; Downing, K. .; Nogales, E. Refined structure of  $\alpha\beta$ -tubulin at 3.5 Å resolution. *J. Mol. Biol.* **2001**, *313*, 1045–1057, doi:10.1006/JMBI.2001.5077.
120. Nogales, E. Preservation of 2-D Crystals of Tubulin for Electron Crystallography. *J. Struct. Biol.* **1995**, *115*, 199–208, doi:10.1006/jsbi.1995.1044.
121. Nogales, E.; Whittaker, M.; Milligan, R. A.; Downing, K. H. High-resolution model of the microtubule. *Cell* **1999**, *96*, 79–88, doi:10.1016/S0092-8674(00)80961-7.
122. Rao, S. T.; Rossmann, M. G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **1973**, *76*, 241–56.

123. Ludueña, R. F. Multiple forms of tubulin: different gene products and covalent modifications. *Int. Rev. Cytol.* **1998**, *178*, 207–75.
124. Roll-Mecak, A. Intrinsically disordered tubulin tails: complex tuners of microtubule functions? *Semin. Cell Dev. Biol.* **2015**, *37*, 11–19, doi:10.1016/j.semcdb.2014.09.026.
125. Wloga, D.; Joachimiak, E.; Fabczak, H. Tubulin Post-Translational Modifications and Microtubule Dynamics. *Int. J. Mol. Sci.* **2017**, *18*, 2207, doi:10.3390/ijms18102207.
126. Verhey, K. J.; Gaertig, J. The Tubulin Code. *Cell Cycle* **2007**, *6*, 2152–2160, doi:10.4161/cc.6.17.4633.
127. Chakraborti, S.; Natarajan, K.; Curiel, J.; Janke, C.; Liu, J. The emerging role of the tubulin code: From the tubulin molecule to neuronal function and disease. *Cytoskeleton* **2016**, *73*, 521–550, doi:10.1002/cm.21290.
128. Gigant, B.; Curmi, P. A.; Martin-Barbey, C.; Charbaut, E.; Lachkar, S.; Lebeau, L.; Siavoshian, S.; Sobel, A.; Knossow, M. The 4 Å X-Ray Structure of a Tubulin:Stathmin-like Domain Complex. *Cell* **2000**, *102*, 809–816, doi:10.1016/S0092-8674(00)00069-6.
129. Charbaut, E.; Curmi, P. A.; Ozon, S.; Lachkar, S.; Redeker, V.; Sobel, A. Stathmin Family Proteins Display Specific Molecular and Tubulin Binding Properties. *J. Biol. Chem.* **2001**, *276*, 16146–16154, doi:10.1074/jbc.M010637200.
130. Curmi, P. A.; Andersen, S. S.; Lachkar, S.; Gavet, O.; Karsenti, E.; Knossow, M.; Sobel, A. The stathmin/tubulin interaction in vitro. *J. Biol. Chem.* **1997**, *272*, 25029–36.
131. Belmont, L. D.; Mitchison, T. J. Identification of a Protein That Interacts with Tubulin Dimers and Increases the Catastrophe Rate of Microtubules. *Cell* **1996**, *84*, 623–631, doi:10.1016/S0092-8674(00)81037-5.
132. Ravelli, R. B. G.; Gigant, B.; Curmi, P. A.; Jourdain, I.; Lachkar, S.; Sobel, A.; Knossow, M. Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain. *Nature* **2004**, *428*, 198–202, doi:10.1038/nature02393.
133. Ranaivoson, F. M.; Gigant, B.; Berritt, S.; Joullié, M.; Knossow, M.; IUCr Structural plasticity of tubulin assembly probed by vinca-domain ligands. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *68*, 927–934, doi:10.1107/S0907444912017143.

134. Prota, A. E.; Bargsten, K.; Northcote, P. T.; Marsh, M.; Altmann, K.-H.; Miller, J. H.; Díaz, J. F.; Steinmetz, M. O. Structural Basis of Microtubule Stabilization by Laulimalide and Peloruside A. *Angew. Chemie Int. Ed.* **2014**, *53*, 1621–1625, doi:10.1002/anie.201307749.
135. Prota, A. E.; Bargsten, K.; Zurwerra, D.; Field, J. J.; Díaz, J. F.; Altmann, K.-H.; Steinmetz, M. O. Molecular mechanism of action of microtubule-stabilizing anticancer agents. *Science* **2013**, *339*, 587–90, doi:10.1126/science.1230582.
136. Giannakakou, P.; Sackett, D.; Fojo, T. Tubulin/microtubules: still a promising target for new chemotherapeutic agents. *J. Natl. Cancer Inst.* **2000**, *92*, 182–3.
137. Hayden, J. H.; Bowser, S. S.; Rieder, C. L. Kinetochore capture astral microtubules during chromosome attachment to the mitotic spindle: direct visualization in live newt lung cells. *J. Cell Biol.* **1990**, *111*, 1039–45.
138. Adapted from Lordjuppiter, W. C. I. Wikimedia Common Image.
139. Jordan, M. A.; Wendell, K.; Gardiner, S.; Derry, W. B.; Copp, H.; Wilson, L. Mitotic block induced in HeLa cells by low concentrations of paclitaxel (Taxol) results in abnormal mitotic exit and apoptotic cell death. *Cancer Res.* **1996**, *56*, 816–25.
140. Jordan, M. A.; Wilson, L. Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer* **2004**, *4*, 253–265, doi:10.1038/nrc1317.
141. Moudi, M.; Go, R.; Yien, C. Y. S.; Nazre, M. Vinca alkaloids. *Int. J. Prev. Med.* **2013**, *4*, 1231–5.
142. Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **1971**, *93*, 2325–7.
143. Jordan, M. A.; Wilson, L. Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer* **2004**, *4*, 253–265, doi:10.1038/nrc1317.
144. Elie-Caille, C.; Severin, F.; Helenius, J.; Howard, J.; Muller, D. J.; Hyman, A. A. Straight GDP-Tubulin Protofilaments Form in the Presence of Taxol. *Curr. Biol.* **2007**, *17*, 1765–1770, doi:10.1016/j.cub.2007.08.063.
145. Dumontet, C.; Jordan, M. A. Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat. Rev. Drug Discov.* **2010**, *9*, 790–803, doi:10.1038/nrd3253.

146. Field, J. J.; Díaz, J. F.; Miller, J. H. The binding sites of microtubule-stabilizing agents. *Chem. Biol.* **2013**, *20*, 301–15, doi:10.1016/j.chembiol.2013.01.014.
147. Rohena, C. C.; Mooberry, S. L. Recent progress with microtubule stabilizers: new compounds, binding modes and cellular activities. *Nat. Prod. Rep.* **2014**, *31*, 335–355, doi:10.1039/C3NP70092E.
148. Schiff, P. B.; Fant, J.; Horwitz, S. B. Promotion of microtubule assembly in vitro by taxol. *Nature* **1979**, *277*, 665–7.
149. Rao, S.; Horwitz, S. B.; Ringel, I. Direct Photoaffinity Labeling of Tubulin With Taxol. *JNCI J. Natl. Cancer Inst.* **1992**, *84*, 785–788, doi:10.1093/jnci/84.10.785.
150. Li, H.; DeRosier, D. J.; Nicholson, W. V; Nogales, E.; Downing, K. H. Microtubule structure at 8 Å resolution. *Structure* **2002**, *10*, 1317–28, doi:10.1016/S0969-2126(02)00827-4.
151. Alushin, G. M.; Lander, G. C.; Kellogg, E. H.; Zhang, R.; Baker, D.; Nogales, E. High-resolution microtubule structures reveal the structural transitions in  $\alpha\beta$ -tubulin upon GTP hydrolysis. *Cell* **2014**, *157*, 1117–29, doi:10.1016/j.cell.2014.03.053.
152. Nogales, E. An electron microscopy journey in the study of microtubule structure and dynamics. *Protein Sci.* **2015**, *24*, 1912–1919, doi:10.1002/pro.2808.
153. Carlomagno, T.; Blommers, M. J. J.; Meiler, J.; Jahnke, W.; Schupp, T.; Petersen, F.; Schinzer, D.; Altmann, K.-H.; Griesinger, C. The High-Resolution Solution Structure of Epothilone A Bound to Tubulin: An Understanding of the Structure–Activity Relationships for a Powerful Class of Antitumor Agents. *Angew. Chemie Int. Ed.* **2003**, *42*, 2511–2515, doi:10.1002/anie.200351276.
154. Canales, A.; Nieto, L.; Rodríguez-Salarichs, J.; Sánchez-Murcia, P. A.; Coderch, C.; Cortés-Cabrera, A.; Paterson, I.; Carlomagno, T.; Gago, F.; Andreu, J. M.; Altmann, K.-H.; Jiménez-Barbero, J.; Díaz, J. F. Molecular Recognition of Epothilones by Microtubules and Tubulin Dimers Revealed by Biochemical and NMR Approaches. *ACS Chem. Biol.* **2014**, *9*, 1033–1043, doi:10.1021/cb400673h.
155. Kellogg, E. H.; Hejab, N. M. A.; Howes, S.; Northcote, P.; Miller, J. H.; Díaz, J. F.; Downing, K. H.; Nogales, E. Insights into the Distinct Mechanisms of Action of Taxane and Non-Taxane Microtubule Stabilizers from Cryo-EM Structures. *J. Mol. Biol.* **2017**, *429*, 633–646, doi:10.1016/j.jmb.2017.01.001.

156. Kavallaris, M. Microtubules and resistance to tubulin-binding agents. *Nat. Rev. Cancer* **2010**, *10*, 194–204, doi:10.1038/nrc2803.
157. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169, doi:10.1093/nar/gkw1099.
158. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, NA-NA, doi:10.1002/prot.22711.
159. Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493–1502, doi:10.1021/ct3000734.
160. Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003**, *24*, 1016–1025, doi:10.1002/jcc.10262.
161. D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. .
162. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–60, doi:10.1016/j.jmgm.2005.12.005.
163. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854, doi:10.1093/bioinformatics/btt055.
164. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 14101, doi:10.1063/1.2408420.
165. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190, doi:10.1063/1.328693.
166. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear



- constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472, doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
167. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593, doi:10.1063/1.470117.
  168. La Sala, G.; Decherchi, S.; De Vivo, M.; Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent. Sci.* **2017**, *3*, 949–960, doi:10.1021/acscentsci.7b00211.
  169. Decherchi, S.; Rocchia, W. A general and Robust Ray-Casting-Based Algorithm for Triangulating Surfaces at the Nanoscale. *PLoS One* **2013**, *8*, e59744, doi:10.1371/journal.pone.0059744.
  170. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174, doi:10.1002/jcc.20035.
  171. Jogalekar, A. S.; Kriel, F. H.; Shi, Q.; Cornett, B.; Cicero, D.; Snyder, J. P. The Discodermolide Hairpin Structure Flows from Conformationally Stable Modular Motifs. *J. Med. Chem.* **2010**, *53*, 155–165, doi:10.1021/jm9015284.
  172. Canales, A.; Matesanz, R.; Gardner, N. M.; Andreu, J. M.; Paterson, I.; Díaz, J. F.; Jiménez-Barbero, J. The Bound Conformation of Microtubule-Stabilizing Agents: NMR Insights into the Bioactive 3D Structure of Discodermolide and Dictyostatin. *Chem. - A Eur. J.* **2008**, *14*, 7557–7569, doi:10.1002/chem.200800039.
  173. Frisch, M. J. Gaussian 03. *Gaussian, Inc* **2003**.
  174. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–8, 27–8.
  175. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613, doi:10.1016/j.cpc.2013.09.018.
  176. Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J. Mol. Biol.* **2002**, *320*, 597–608, doi:10.1016/S0022-2836(02)00470-9.

177. Schrodinger LLC The PyMOL Molecular Graphics System, Version 1.8 2015.
178. Janert, P. K. *Gnuplot in action : understanding data with graphs*; ISBN 1633430189.
179. Connolly, M. L.; IUCr Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558, doi:10.1107/S0021889883010985.
180. Richards, F. M. Areas, Volumes, Packing, and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176, doi:10.1146/annurev.bb.06.060177.001055.
181. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749, doi:10.1021/jm0306430.
182. Canales, A.; Rodríguez-Salarichs, J.; Trigili, C.; Nieto, L.; Coderch, C.; Andreu, J. M.; Paterson, I.; Jiménez-Barbero, J.; Díaz, J. F. Insights into the Interaction of Discodermolide and Docetaxel with Tubulin. Mapping the Binding Sites of Microtubule-Stabilizing Agents by Using an Integrated NMR and Computational Approach. *ACS Chem. Biol.* **2011**, *6*, 789–799, doi:10.1021/cb200099u.
183. Shujun Xia, ‡; Craig S. Kenesky, §; Paul V. Rucker, §; Amos B. Smith, I.; George A. Orr, ‡ and; Susan Band Horwitz\*, ‡ A Photoaffinity Analogue of Discodermolide Specifically Labels a Peptide in  $\beta$ -Tubulin†. **2006**, doi:10.1021/BI060497A.
184. Khrapunovich-Baine, M.; Menon, V.; Verdier-Pinard, P.; Smith, A. B.; Angeletti, R. H.; Fiser, A.; Horwitz, S. B.; Xiao, H. Distinct Pose of Discodermolide in Taxol Binding Pocket Drives a Complementary Mode of Microtubule Stabilization. *Biochemistry* **2009**, *48*, 11664–11677, doi:10.1021/bi901351q.
185. Protá, A. E.; Bargsten, K.; Redondo-Horcajo, M.; Smith, A. B.; Yang, C.-P. H.; McDaid, H. M.; Paterson, I.; Horwitz, S. B.; Fernando Díaz, J.; Steinmetz, M. O. Structural Basis of Microtubule Stabilization by Discodermolide. *ChemBioChem* **2017**, *18*, 905–909, doi:10.1002/cbic.201600696.
186. Yang, C.-P.; Horwitz, S. Taxol®: The First Microtubule Stabilizing Agent. *Int. J. Mol. Sci.* **2017**, *18*, 1733, doi:10.3390/ijms18081733.